

# Enzyklopädie der Psychologie

# ENZYKLOPÄDIE DER PSYCHOLOGIE

In Verbindung mit der  
Deutschen Gesellschaft für Psychologie

herausgegeben von

Prof. Dr. Carl F. Graumann, Heidelberg  
Prof. Dr. Theo Herrmann, Mannheim  
Prof. Dr. Hans Hörmann, Bochum  
Prof. Dr. Martin Irle, Mannheim  
Prof. Dr. Dr. h.c. Hans Thomae, Bonn  
Prof. Dr. Franz E. Weinert, München

Themenbereich B

Methodologie und Methoden

Serie I

Forschungsmethoden der Psychologie

Band 4

Strukturierung und Reduzierung von Daten



Verlag für Psychologie · Dr. C. J. Hogrefe  
Göttingen · Toronto · Zürich

# Strukturierung und Reduzierung von Daten

Herausgegeben von

Prof. Dr. Jürgen Bredenkamp, Trier  
und Prof. Dr. Hubert Feger, Hamburg



Verlag für Psychologie · Dr. C. J. Hogrefe  
Göttingen · Toronto · Zürich

© by Verlag für Psychologie Dr. C.J. Hogrefe, Göttingen 1983  
Alle Rechte, insbesondere das der Übersetzung in fremde Sprachen, vorbehalten

Gesamtherstellung: Allgäuer Zeitungsverlag GmbH, 8960 Kempten (Allgäu)  
Printed in Germany

ISBN 3 8017 0514 5

## Autorenverzeichnis

*Dr. Klaus Haagen*

Lehrstuhl für Angewandte  
Stochastik der  
Universität München  
Akademiestraße 1/IV  
D - 8000 München 22

*Prof. Dr. Walter Oberhofer*

Institut für Volkswirtschafts-  
lehre und Ökonometrie  
der Universität Regensburg  
Universitätsstraße 31  
D - 8400 Regensburg

*Prof. Dr. Joachim Krauth*

Psychologisches Institut  
der Universität Düsseldorf  
Lehrstuhl für Psychologie IV  
Universitätsstraße 1  
D - 4000 Düsseldorf 1

*Dr. Hartmut A. Oldenbürger*

Seminar für Wirtschaftspädagogik  
der Universität Göttingen  
Goßlerstraße 12  
D - 3400 Göttingen

*Prof. Dr. Helfried Moosbrugger*

Fachbereich 5 - Psychologie der  
Johann-Wolfgang-Goethe-Universität  
Mertonstraße 17  
D - 6000 Frankfurt/Main

*Dr. Werner Schubö*

Institut für Psychologie  
der Universität München  
- Allgemeine Psychologie -  
Friedrichstraße 22  
D - 8000 München 40

*Dr. Rolf Steyer*

Fachbereich I - Psychologie  
der Universität Trier  
Schneidershof  
D - 5500 Trier

# Inhaltsverzeichnis

## 1. Kapitel: Modelle zur Beschreibung statistischer Zusammenhänge in der psychologischen Forschung. Von Helfried Moosbrugger

1. Einführung und Überblick . . . . .	1
2. Modelle mit manifesten Variablen . . . . .	5
2.1 Einleitung und Überblick . . . . .	5
2.2 Eine formale Theorie zur Beschreibung statistischer Zusammenhänge . . . . .	6
2.2.1 Die Grundannahme . . . . .	6
2.2.2 Die Residualvariable . . . . .	6
2.2.3 Kovarianzmodellgleichungen und Parameteridentifikation . . . . .	7
2.2.4 Determinierte Varianz, multiple Korrelation und Determinationskoeffizient . . . . .	9
2.2.5 Multivariate Verallgemeinerung . . . . .	9
2.2.6 Zusammenfassende Bemerkungen . . . . .	10
2.3 Anwendungen der formalen Theorie . . . . .	10
2.3.1 Regressions- und Korrelationsanalyse . . . . .	11
2.3.2 Varianzanalyse . . . . .	14
2.3.3 Diskriminanzanalyse . . . . .	21
2.3.4 Kontingenzanalyse . . . . .	23
3. Modelle mit latenten Variablen . . . . .	25
3.1 Einleitung . . . . .	25
3.2 Eine allgemeine Theorie latenter Variablen . . . . .	27
3.2.1 Die Grundannahmen . . . . .	27
3.2.1.1 Bedingte Unabhängigkeit . . . . .	27
3.2.1.2 Variablencharakteristische Funktion (VC-Funktion) . . . . .	28
3.2.2 Die Residualvariablen . . . . .	30
3.2.3 Kovarianzmodellgleichungen und Identifikation . . . . .	32
3.2.4 Beispiel . . . . .	34
3.2.5 Modell paralleler Variablen . . . . .	35
3.2.6. Zusammenfassende Bemerkungen . . . . .	36

<b>3.3 Anwendungen der allgemeinen Theorie latenter Variablen . . . . .</b>	<b>36</b>
3.3.1 Faktorenanalyse . . . . .	36
3.3.2 „Linear traceline model“ . . . . .	38
3.3.3 „Latent profile“ und „latent class model“ . . . . .	40
3.3.4 Logistische Modelle . . . . .	43
3.3.5 Klassisches latent-additives Testmodell . . . . .	45
<b>3.4 Zusammenfassende Bemerkungen . . . . .</b>	<b>47</b>
<b>4. Anhänge . . . . .</b>	<b>47</b>
<b>4.1 Anhang A:</b>	
Regeln für Erwartungswerte (vgl. Hays, 1973, S. 871 ff.) . . . . .	47
<b>4.2 Anhang B:</b>	
Regeln für Varianzen und Kovarianzen . . . . .	48
<b>4.3 Anhang C:</b>	
Regeln für bedingte Erwartungen . . . . .	50

## 2. Kapitel: Modelle zur kausalen Erklärung statistischer Zusammenhänge. Von Rolf Steyer

<b>1. Einführung . . . . .</b>	<b>59</b>
1.1 Zur Bedeutsamkeit kausaler Abhängigkeit . . . . .	59
1.2 Zum Forschungsstand . . . . .	61
1.3 Überblick . . . . .	66
<b>2. Münzen und Elektromagnet . . . . .</b>	<b>67</b>
2.1 Einleitende Bemerkungen . . . . .	67
2.2 Beschreibung des Beispiels . . . . .	68
2.3 Abhängigkeit der ersten von der zweiten Münzvariablen . . . . .	72
2.4 Abhängigkeit der Münz- von der Magnetvariablen . . . . .	73
2.5 Das Problem . . . . .	76
2.6 Zusammenfassende Bemerkungen . . . . .	76
<b>3. Interne Validität . . . . .</b>	<b>76</b>
3.1 Einleitende Bemerkungen . . . . .	76
3.2 Grundideen . . . . .	77
3.3 Fälle, in denen keine interne Validität besteht . . . . .	80
3.4 Fälle, in denen möglicherweise interne Validität besteht . . . . .	83
3.5 Zusammenfassende Bemerkungen . . . . .	87

4. Einfache kausale reglineare Abhangigkeit . . . . .	88
4.1 Einleitende Bemerkungen . . . . .	88
4.2 Vorgeordnetheit . . . . .	88
4.3 Invarianz . . . . .	92
4.4 Definition . . . . .	95
4.5 Beispiel: Munzen und Elektromagnet (1. Fortsetzung) . . . . .	97
4.6 Beispiel: Drogen und Aktivierung . . . . .	98
4.7 Zusammenfassende Bemerkungen . . . . .	101
5. Eigenschaften einfacher kausalerreglinearer Abhangigkeit . . . . .	101
5.1 Einleitende Bemerkungen . . . . .	101
5.2 Unkonfundiertheit . . . . .	102
5.3 Beispiel: Munzen und Elektromagnet (1. Fortsetzung) . . . . .	103
5.4 Vollstandige Abhangigkeit . . . . .	105
5.5 Faktische Konstanthaltung . . . . .	108
5.6 Unabhangigkeit . . . . .	110
5.7 Randomisierung und Parallelisierung . . . . .	113
5.8 Zusammenfassende Bemerkungen . . . . .	115
6. Munze und Elektromagnet mit zwei Schaltern . . . . .	116
6.1 Einleitende Bemerkungen . . . . .	116
6.2 Beschreibung des Beispiels . . . . .	117
6.3 Reglineare Abhangigkeit . . . . .	119
6.4 Logitlineare Abhangigkeit . . . . .	122
6.5 Zusammenfassende Bemerkungen . . . . .	123
7. Externe Validitat . . . . .	123
7.1 Einleitende Bemerkungen . . . . .	123
7.2 Situationsvaliditat . . . . .	124
7.3 Populationsvaliditat . . . . .	127
7.4 Vergleiche der externen Validitat . . . . .	131
7.5 Zusammenfassende Bemerkungen . . . . .	133
8. Ausblick . . . . .	133
8.1 Mehrvariablenmodelle . . . . .	133
8.2 Beschreibende und erklarende reglineare Modelle . . . . .	137
9. Weiterfuhrende Literatur . . . . .	140
Anhang . . . . .	140
A.1 Einleitende Bemerkungen . . . . .	140
A.2 Erwartungswert . . . . .	140

A.3 Varianz und Kovarianz . . . . .	141
A.4 Bedingter Erwartungswert . . . . .	142
A.5 Bedingte Erwartung . . . . .	143

### 3. Kapitel: Uni- und multivariate Varianzanalyse mit festen Parametern. Von Helfried Moosbrugger und Rolf Steyer

1. Einführung und Überblick . . . . .	154
2. Multivariate lineare Modelle mit festen Parametern . . . . .	155
2.1 Einleitung . . . . .	155
2.2 Die grundlegenden Modellvorstellungen . . . . .	156
2.3 Stichprobenmodelle . . . . .	161
2.4 Zusammenfassende Bemerkungen . . . . .	164
3. Hypothesenformulierung verschiedenen Designs . . . . .	165
3.1 Einleitung . . . . .	165
3.2 Das Zellenmittelwertmodell . . . . .	165
3.3 Die multivariate allgemeine lineare Hypothese . . . . .	168
3.4 Gekreuzte Faktoren über den unabhängigen Variablen . . . . .	169
3.5 Gekreuzte Faktoren über den abhängigen Variablen . . . . .	173
3.6 Hierarchische Faktoren über den unabhängigen Variablen . . . . .	177
3.7 Hierarchische Faktoren über den abhängigen Variablen . . . . .	179
3.8 Lateinisches Quadrat über den unabhängigen Variablen . . . . .	182
3.9 Lateinisches Quadrat über den abhängigen Variablen . . . . .	183
3.10 Zusammenfassende Bemerkungen . . . . .	185
4. Parameterschätzung . . . . .	186
4.1 Einleitung . . . . .	186
4.2 Kriterium der kleinsten Quadrate . . . . .	187
4.3 Kriterium der kleinsten Quadrate unter Nebenbedingungen . . . . .	188
4.4 Maximum-Likelihood-Kriterium . . . . .	190
4.5 Erwartungswerte- und Kovarianzmatrix der Parametervektoren $\beta_p$ und $\psi_k$ . . . . .	190
4.6 Zusammenfassende Bemerkungen . . . . .	192
5. Hypothesenbewertung . . . . .	192
5.1 Einleitung . . . . .	192
5.2 Wilks' Lambda-Kriterium . . . . .	194
5.3 Roy's Eigenwert-Kriterium . . . . .	195

5.4 Hotelling - Lawley Spur Kriterium . . . . .	195
5.5 Pillai-Bartlett Spur Kriterium . . . . .	196
5.6 Einfache Konfidenzintervalle . . . . .	196
5.6.1 Konfidenzintervall für einen Parameter $\beta_{qp}$ . . . . .	196
5.6.2 Konfidenzintervall für Realisationen einer abhängigen Variablen $y_p$ . . . . .	198
5.7 Praktische Signifikanz . . . . .	199
5.8 Zusammenfassende Bemerkungen . . . . .	199

## 4. Kapitel: Regressions- und kanonische Analyse.

Von Werner Schubö, Klaus Haagen und Walter Oberhofer

1. Regressionsanalyse . . . . .	207
1.1 Beschreibende lineare Regression . . . . .	209
1.2 Das allgemeine regressionsanalytische Modell . . . . .	214
1.3 Die Schätzung der Parameter im allgemeinen regressionsanalytischen Modell . . . . .	217
1.4 Prognose im allgemeinen regressionsanalytischen Modell . . . . .	219
1.5 Statistische Tests im klassischen regressionsanalytischen Modell . . . . .	223
1.6 Ridge-Regression . . . . .	230
1.7 Klassisches korrelationsanalytisches Modell und multiple Korrelation . . . . .	233
1.8 Modelle mit Fehlern in den Prädiktoren . . . . .	237
1.9 Zeitreihenanalyse im allgemeinen regressionsanalytischen Modell . . . . .	243
1.10 Suppression und Kollinearität . . . . .	249
1.11 Schrittweise Regression . . . . .	257
1.12 Teststärke . . . . .	261
1.12.1 Teststärke für die Prüfung der multiplen Regression . . . . .	263
1.12.2 Bestimmendes erforderlichen Stichprobenumfangs N . . . . .	263
1.12.3 Die erforderliche Populationskorrelation R . . . . .	267
1.12.4 Die höchstens sinnvolle Prädiktorenzahl K . . . . .	267
1.12.5 Teststärke für die übrigen Signifikanztests bei der Regressionsanalyse . . . . .	267
2. Kanonische Korrelation . . . . .	270
2.1 Einführung . . . . .	270
2.2 Das Modell der kanonischen Korrelation für zwei Variablenmengen mit zufälligen Größen . . . . .	271
2.3 Schätzung der kanonischen Korrelationen und der Koeffizientenvektoren der kanonischen Variablen . . . . .	276

2.4 Test zur Bestimmung der Anzahl der kanonischen Variablen . . . . .	277
2.5 Extraktions- und Redundanzmaße . . . . .	278
2.6 Verallgemeinerung der kanonischen Korrelation auf mehr als zwei Variablenmengen . . . . .	281

## 5. Kapitel: Diskriminanzanalyse. Von Joachim Krauth

1. Einführung . . . . .	293
1.1 Problemstellung . . . . .	293
1.2 Entstehungsgeschichte . . . . .	295
1.3 Übersichtsarbeiten . . . . .	295
2. Grundlagen . . . . .	295
2.1 Lineare Diskriminanzfunktion . . . . .	295
2.2 Bayes-Ansatz . . . . .	300
2.3 Fehlerraten . . . . .	301
2.4 Minimaxprinzip . . . . .	303
2.5 Diskriminanzanalyse unter Nebenbedingungen und Kosten der Fehlklassifikation . . . . .	304
2.6 Quadratische Diskrimination . . . . .	305
2.7 Zusammenhang zwischen Diskrimination und Regression . . . . .	307
2.8 Verfahren für mehrere Populationen . . . . .	308
2.8.1 Multiple Diskriminanzanalyse . . . . .	308
2.8.2 Minimierung des erwarteten Verlustes . . . . .	310
2.8.3 Distanzmaße . . . . .	311
2.8.4 Andere Verfahren . . . . .	311
2.8.5 Methodenvergleich . . . . .	312
2.9 Logistische Diskrimination . . . . .	313
2.10 Kovariante Diskriminanzanalyse . . . . .	314
2.11 Sequentielle Diskrimination . . . . .	315
2.12 Zeitreihen . . . . .	317
2.13 Variablenauswahl . . . . .	318
3. Inferenzstatistik . . . . .	321
3.1 Signifikanztests . . . . .	321
3.2 Schätzungen . . . . .	322
4. Robustheit . . . . .	323
4.1 Lineare Diskriminanzfunktion . . . . .	323

4.2 Quadratische Diskriminanzfunktion . . . . .	324
4.3 Robuste Diskriminanzfunktionen . . . . .	324
5. Nichtparametrische Verfahren . . . . .	325
5.1 Nichtparametrische Zuordnungsregeln . . . . .	325
5.2 Variablenauswahl . . . . .	327
5.3 Schätzungen der Fehlerrate . . . . .	328
6. Analyse qualitativer und diskreter Daten . . . . .	328
6.1 Verteilungsmodelle . . . . .	328
6.1.1 Volles Multinomialmodell . . . . .	328
6.1.2 Modelle bei multivariaten binären Items . . . . .	329
6.2 Nichtparametrische Verfahren bei qualitativen Daten . . . . .	330
6.3 Gleichzeitiges Vorliegen diskreter und stetiger Variablen . . . . .	331
6.4 Variablenauswahl . . . . .	331

## 6. Kapitel: Latente Strukturanalyse. Von Joachim Krauth

1. Einführung . . . . .	351
2. Grundbegriffe der latenten Strukturanalyse . . . . .	353
3. Allgemeines Vorgehen bei der latenten Strukturanalyse . . . . .	357
4. Modelle der latenten Strukturanalyse . . . . .	359
4.1 Allgemeines Modell . . . . .	359
4.2 Existenzproblem . . . . .	361
4.3 Identifikationsproblem . . . . .	361
4.4 Strukturproblem . . . . .	362
4.5 Latentes Klassenmodell . . . . .	362
4.6 Latentes Polynommodell . . . . .	367
4.7 Lokalisiertes Klassenmodell . . . . .	368
4.8 Latentes Inhaltsmodell . . . . .	369
4.9 Latentes Distanzmodell . . . . .	370
4.10 Testtheoretisches Modell . . . . .	370
4.11 Latentes Profilmodell . . . . .	371
4.12 Andere Modelle . . . . .	372
5. Statistische Fragestellungen . . . . .	373
5.1 Parameterschätzung . . . . .	373

5.1.1 Einführung . . . . .	373
5.1.2 Algebraische Verfahren . . . . .	373
5.1.3 Faktorisierungsmethoden . . . . .	375
5.1.4 Maximum-Likelihood-Schätzungen . . . . .	375
5.1.5 Aufteilungsmethoden . . . . .	377
5.1.6 Andere Schätzmethoden . . . . .	378
5.1.7 Programme und Algorithmen . . . . .	378
5.1.8 Probleme beim Schätzen . . . . .	379
5.2 Signifikanztests . . . . .	380
6. Schätzung der latenten Variablen . . . . .	380
7. Vergleich mit anderen Verfahren . . . . .	381
7.1 Vergleich mit der Faktorenanalyse . . . . .	381
7.2 Vergleich mit der Guttman-Skalierung . . . . .	382
8. Anwendungen . . . . .	383

## 7. Kapitel: Clusteranalyse. Von Hartmut-A. Oldenbürger

1. Zur Entwicklung der Literatur . . . . .	390
2. Zur Datenerhebung und Datenstruktur . . . . .	392
3. Problemstellungen und Verfahren . . . . .	395
3.1 Untermengenauswahl . . . . .	395
3.2 Mengenerlegung . . . . .	398
3.3 Hierarchische Clusteranalysen . . . . .	404
3.3.1 Einordnung und Charakteristik . . . . .	404
3.3.2 Agglomerative Verfahren . . . . .	407
3.3.3 Subdivisive Verfahren . . . . .	412
3.3.4 Evaluation, Anwendung und Weiterentwicklungen . . . . .	414
3.4 Baumrepräsentationen und hybride Modelle . . . . .	419
3.5 Überlappende Gruppierung . . . . .	421
3.6 Cluster in Datenmatrizen . . . . .	424
4. Diskussion und Ausblick . . . . .	425

## 8. Kapitel: Typenanalyse. Von Joachim Krauth

1. Einführung . . . . .	440
2. Die Musteranalyse von McQuitty . . . . .	442
3. Die Konfigurationsfrequenzanalyse von Lienert . . . . .	460
4. Andere typenanalytische Ansätze . . . . .	470
4.1 <i>Die Musterähnlichkeitsanalyse von Cattell</i> . . . . .	470
4.2 <i>Die Profildistanzanalyse von Sawrey, Keller und Conger</i> . . . . .	472
4.3 <i>Die Übereinstimmungsanalyse von Gengerelli</i> . . . . .	473
4.4 <i>Die Ähnlichkeitsanalyse nach Lorr und McNair</i> . . . . .	475
4.5 <i>Die Gruppierungsanalyse von Friedman und Rubin</i> . . . . .	477
4.6 <i>Der informationstheoretische Ansatz von Wallure und Boulton</i> . . . . .	478
4.7 <i>Die lineare Typenanalyse von Overall und Klett</i> . . . . .	478
4.8 <i>Dieparametrische Mischungsanalyse von Wolfe und Fleiss</i> . . . . .	480
4.9 <i>Die Ähnlichkeitspartialisierungsmethode von Bolz</i> . . . . .	481
<b>Autoren-Register</b> . . . . .	497
<b>Sach-Register</b> . . . . .	509

## 1. Kapitel

# Modelle zur Beschreibung statistischer Zusammenhänge in der psychologischen Forschung

*Helfried Moosbrugger*

## 1. Einführung und Überblick

In diesem Kapitel geben wir eine Einführung in Bereiche der angewandten Statistik, die bereits um die Jahrhundertwende betreten wurden (siehe z.B. Pearson 1896, Spearman 1904), aber bis heute in einer intensiven Entwicklung stehen, deren Ende noch nicht abzusehen ist. Wir behandeln in Form eines integrierenden Überblicks Theorie und Anwendung von Modellen zur Beschreibung statistischer Zusammenhänge.

Beobachtung und Beschreibung von Zusammenhängen haben in der psychologischen Forschung zumindest zwei Funktionen: zum einen entsteht dadurch ein Anlaß, induktiv die Entwicklung inhaltlicher Theorien zu betreiben, weil beobachtete statistische Zusammenhänge einer theoretischen Erklärung bedürfen; zum anderen sind beobachtete statistische Zusammenhänge gewissermaßen die empirischen Gegebenheiten, an denen deduktiv die Theorien auf ihre Gültigkeit hin überprüft werden. Oft sind dabei die empirischen Beobachtungen selbst schon auf theoretische Vorannahmen gegründet: Stellt man beispielsweise bei einer empirischen Untersuchung eine Korrelation zwischen „Intelligenz“ und „Schulleistung“ fest, so hat man vorher implizit im Sinne von Spearman (1904, 1926) die theoretische Entscheidung getroffen, daß „Intelligenz“ eine einheitliche latente Variable sei, die mit einem psychometrischen Test gemessen werden kann. Dieses Vorgehen ist keineswegs unstrittig, weil von anderen Autoren (z.B. Thurstone 1938, Guilford 1967) in Frage gestellt wird, daß es sich bei Intelligenz um eine einheitliche Eigenschaft handle, welche mit dem Intelligenzquotienten als Kennzahl darstellbar ist. Entsprechendes gilt auch für „Schulleistung“. Unter diesem Blickwinkel müssen selbst „empirische“ beobachtete Zusammenhänge bereits als „theoriebehaftet“ (vgl. dazu auch Holzkamp, 1972, S. 81) angesehen werden.

Während wir uns im folgenden Abschnitt mit Modellen befassen, welche die statistischen Zusammenhänge manifester, d.h. direkt beobachtbarer Variablen beschreiben, beschäftigen wir uns im darauffolgenden Abschnitt mit Situationen, in denen beobachtbare Zusammenhänge von manifesten Variablen auf nicht direkt beobachtbare „latente“ Variablen zurückgeführt werden. Solche latenten Variablen sind häufig Begriffe, welche in psychologischen Theorien verwendet werden. „Intelligenz“ z.B. kann nicht als direkt beobachtbar aufgefaßt werden. Beobachtbar sind vielmehr nur Verhaltensweisen, die auf unterschiedliche Ausprägung der Intelligenz schließen lassen. Ebenso verhält es sich mit anderen „Persönlichkeitseigenschaften“. Daher geben wir im dritten Abschnitt eine Einführung in die wesentlichen Grundgedanken der Theorie latenter Variablen, welche insofern die formale Grundlage vieler psychologischer Begriffe bildet, als sie angibt, wie latente Eigenschaften mit beobachtbarem Verhalten verknüpft werden können. Exemplarisch behandeln wir den Fall einer latenten Variablen, welche die Abhängigkeiten zwischen den beobachtbaren Variablen erklärt in dem Sinn, daß deren Abhängigkeiten verschwinden, wenn die gemeinsame latente Variable auf einem festen Wert gehalten wird. Beobachtete Variablen mit dieser Eigenschaft heißen kongenerisch (Jöreskog, 1971, Steyer & Moosbrugger, 1979).

Die formale Gemeinsamkeit aller in diesem Kapitel behandelten statistischen Modelle besteht darin, daß sie alle mit dem Konzept der bedingten Erwartung formuliert werden können, für die im Anhang C einige Rechenregeln zusammengestellt sind. Die bedingte Erwartung  $E(y | x)$  beispielsweise ist eine stochastische Variable oder Zufallsvariable<sup>1)</sup>, deren Werte  $E(y | x=x)$  die Erwartungswerte der abhängigen stochastischen Variablen  $y$  unter der Bedingung sind, daß die „bedingende“ Variable  $x$  einen bestimmten Wert  $x$  angenommen hat. Im Grunde beschreiben die hier behandelten Modelle immer nur in einer mathematischen Formulierung unsere Vorstellung, welche Werte der Variablen  $y$  bei gegebenen Werten der Variablen  $x$  zu erwarten sind. Diese einfache Grundidee kann zu recht komplizierten Modellen mit mehreren  $x$ - und  $y$ -Variablen ausgebaut werden.

Im Abschnitt 2 dieses Kapitels stellen wir statistische Modelle mit manifesten Variablen dar, die alle mit der Gleichung

$$(1.1) \quad E(y | x_1, x_2, \dots, x_Q) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_Q x_Q$$

formuliert werden können, wobei  $y$  sowie die  $x_q, q \in \{1, 2, \dots, Q\}$  stochastische Variablen und die Parameter  $\beta_0$  und  $\beta_q$  reellwertige Konstanten sind. Statt

---

<sup>1)</sup> Wir gebrauchen „stochastische Variable“ synonym zu „Zufallsvariable“ (auf dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$ ), worunter man eine  $\mathcal{A}$ -meßbare Abbildung von  $\Omega$  nach IR, der Menge der reellen Zahlen versteht, wobei  $\Omega$  die Menge der Elementarergebnisse,  $\mathcal{A}$  die (Ereignis-) $\sigma$ -Algebra und  $P$  ein Wahrscheinlichkeitsmaß ist (vgl. Bauer, 1974, s. 137).

$E(y|x_1, x_2, \dots, x_Q)$  schreiben wir auch  $E(y|x)$  mit dem Q-varianten stochastischen Vektor<sup>2)</sup>  $x = (x_1, x_2, \dots, x_Q)$ .

Wie wir zeigen werden, liegt Gleichung 1.1 einer Reihe von Analyseverfahren zugrunde, welche sich nur hinsichtlich der Skalenqualität der beteiligten Variablen unterscheiden: In der Regressionsanalyse sind sowohl  $y$  als auch die  $x_q$  quantitative Variablen, und  $E(y|x)$  ist somit die Regressionsgerade, -ebene oder -hyperebene, je nachdem, wie viele Variablen in  $x$  zusammengefaßt sind. In der Varianzanalyse ist  $y$  quantitativ, aber die  $x_q$  sind qualitative zweistufige Indikatorvariablen, welche die Gruppenzugehörigkeit anzeigen; die Werte von  $E(y|x)$  sind dann Erwartungswerte der abhängigen Variablen in den Gruppen. In der Diskriminanzanalyse mit zwei Gruppen ist  $y$  eine Indikatorvariable für die Gruppenzugehörigkeit, wohingegen die  $x_q$  quantitativ, aber auf ein bestimmtes Intervall beschränkt sind, wenn man auch dort die Gleichung 1.1 zugrunde legt. Kodiert man die Zugehörigkeit zu den beiden Gruppen mit Null und Eins, so hat  $E(y|x)$  die bedingten Wahrscheinlichkeiten  $P(y=1|x=x)$  als Werte, nämlich die Wahrscheinlichkeiten bei gegebenen Werten der Variablen  $x_q$ , der Gruppe „1“ anzugehören.

In der Kontingenzanalyse schließlich sind sowohl  $y$  als auch  $x_q$  qualitative Indikatorvariablen, und die Kodierung von  $y$  mit Null und Eins führt wieder dazu, daß  $E(y|x)$  bedingte Wahrscheinlichkeiten als Werte enthält, aber diesmal unter einer Gruppenzugehörigkeit und nicht, wie bei der Diskriminanzanalyse, unter der Ausprägung quantitativer Variablen  $x_q$ .

Man kann schon hier erkennen, daß alle genannten Verfahren auf der gleichen Modellannahme basieren und sich lediglich durch das Skalenniveau der beteiligten Variablen unterscheiden.

Im Abschnitt 3 behandeln wir ebenfalls Modelle, denen eine Gleichung der Form der Gleichung 1.1 zugrundeliegt, jedoch handelt es sich dort bei der unabhängigen bedingenden Variablen nicht um eine direkt beobachtbare, sondern um eine latente Variable, die wir mit dem kleinen griechischen Buchstaben  $\xi$  bezeichnen. Der Übersichtlichkeit halber beschränken wir uns in jenem dritten Abschnitt auf eine latente Variable. Dafür ist es aber notwendig, zugeleich mehrere abhängige Variablen  $y_p$  zu betrachten. Den meisten Modellen jenes Abschnitts liegt die Gleichung

$$(1.2) \quad E(y_p|\xi) = \lambda_{p0} + \lambda_p \xi \quad p \in \{1, 2, \dots, P\}$$

zugrunde, wobei  $\lambda_{p0}$  wiederum eine reellwertige Skalenkonstante ist und  $\lambda_p$  ein reellwertiger Parameter, der ebenso wie die Konstanten  $\beta_q$  (siehe Gleichung 1.1) die Enge des statistischen Zusammenhangs zwischen den betreffenden Variablen angibt.

---

<sup>2)</sup> Ein Q-varianter stochastischer Vektor (oder Zufallsvektor) ist eine d-meßbare Abbildung von  $\Omega$  nach  $IR^Q$  (vgl. auch Fußnote 1).

Nach der Entwicklung einer allgemeinen Theorie latenter Variablen werden wir zeigen, daß auch Gleichung 1.2 verschiedenen, eng miteinander verwandten Modellen mit latenten Variablen zugrunde liegt, welche sich nur dadurch unterscheiden, daß die beteiligten Variablen quantitativ bzw. qualitativ sind: In der Faktorenanalyse mit einem Faktor, welche formal als Analogon zur Regressionsanalyse aufgefaßt werden kann, sind sowohl die  $y_p$  als auch  $\xi$  quantitativ, und  $E(y_p|\xi)$  entspricht einer Regressionsgeraden. Sind die  $y_p$  wieder Indikatorvariablen für jeweils zwei Gruppen, wohingegen  $\xi$  quantitativ, aber auf ein bestimmtes Intervall beschränkt ist, so erhalten wir Lazarsfeld's „*linear traceline model*“, wobei  $E(y_p|\xi)$  die „*traceline*“ bezeichnet und wieder bedingte Wahrscheinlichkeiten  $P(y_p=1|\xi=\xi)$  als Werte aufweist, wenn man als Kodierung für die  $y_p$  Null und Eins wählt. In formaler Hinsicht entspricht dieses Modell daher der Diskriminanzanalyse. Sind die  $y_p$  quantitativ und ist  $\xi$  eine Variable, welche die Zugehörigkeit zu einer von zwei latenten Gruppen anzeigen, so erhalten wir Gibsons „*latent profile model*“. Die bedingte Erwartung  $E(y_p|\xi)$  enthält dabei die Erwartungswerte der  $y_p$  innerhalb der jeweiligen latenten Gruppe. Formal ist dieses Modell analog zur Varianzanalyse. Sind schließlich alle beteiligten Variablen qualitativ mit der Kodierung Null und Eins, so finden wir, daß Gleichung 1.2 das „*latent class model*“ von Lazarsfeld beschreibt, und  $E(y_p|\xi)$  hat wieder die bedingten Wahrscheinlichkeiten  $P(y_p=1|\xi=\xi)$  als Werte. Dieses Modell entspricht also in formaler Hinsicht die Kontingenzanalyse. Auch „*logistische Testmodelle*“ wie z.B. das „*Rasch-Modell*“, und das „*klassische latent-additive Testmodell*“ von Moosbrugger & Müller, welche keine unmittelbare Entsprechung bei den Modellen mit manifesten Variablen haben, werden wir als Anwendungsfälle der Theorie latenter Variablen referieren.

Bei allen Gemeinsamkeiten, die in formaler Hinsicht zwischen den Modellen der zweiten und denjenigen des dritten Abschnitts bestehen, darf man jedoch nicht die Unterschiedlichkeit übersehen, die in ihren Zielen bestehen. Während die Modelle des zweiten Abschnitts den Zweck haben, statistische Zusammenhänge zu beschreiben, dienen die Modelle des dritten Abschnitts gewissermaßen der Bildung von psychologischen Begriffen. Die jeweiligen latenten Variablen  $\xi$  lassen sich nämlich als psychologische Konstrukte (s. z.B. Herrmann, 1973 oder MacCorquodale & Meehl, 1948) interpretieren.

Die Modelle latenter Variablen können als eine Art der Erklärung von statistischen Zusammenhängen aufgefaßt werden, da die Zusammenhänge zwischen den abhängigen Variablen  $y_p$ , sofern sie „kongenerisch“ sind, bei jeder Ausprägung von  $\xi$  verschwinden. In diesem Sinn kann man sagen, daß die latente Variable  $\xi$  dem Modell nach die statistischen Abhängigkeiten zwischen den beobachtbaren Variablen  $y_p$  erklärt. Eine grundlegende Behandlung von Modellen zur Erklärung statistischer Zusammenhänge nimmt Steyer (1982) in diesem Band vor.

## 2. Modelle mit manifesten Variablen

### 2.1 Einleitung und Überblick

Im vorliegenden Abschnitt 2 behandeln wir die Regressions-, Varianz-, Diskriminanz- und Kontingenzanalyse. Dabei zeigen wir, daß die Analyseverfahren in formaler Hinsicht grundlegende Gemeinsamkeiten aufweisen. Die verschiedenen Namen dieser Verfahren sind eher historisch zu verstehen, ebenso wie ihre getrennte Behandlung in den Lehrbüchern der angewandten Statistik. Die Worte „Varianz-“, „Regressionsanalyse“ etc. sind eher als Kennzeichnung für die Fragestellungen aufzufassen, bei denen diese Verfahren angewandt werden. Alle in diesem Abschnitt behandelten Modelle präzisieren Vorstellungen über statistische Zusammenhänge direkt beobachtbarer oder „manifester“ Variablen.

Statistische Zusammenhänge können erstens zur Prädiktion verwendet werden, was ein vorrangiges Ziel in der „Regressionanalyse“ ist. Nachdem man in einer Stichprobe das „Zusammenauftreten“ zweier Variablen beobachtet hat, kann man in anderen Stichproben den Wert der einen Variablen aus dem Wert der anderen vorhersagen. Die Stärke statistischer Zusammenhänge kann zweitens aber auch für sich genommen von Interesse sein. Sie wird in der Korrelationsanalyse untersucht. Drittens sind statistische Zusammenhänge zwischen Gruppen einerseits - welche als qualitative Variablen kodierbar sind - und quantitativen Variablen andererseits ein Ausdruck von Unterschieden zwischen den Gruppen hinsichtlich der Mittelwerte der quantitativen Variablen. Die Untersuchung von Mittelwerteunterschieden ist unter den Begriffen „*t*-Test“, „Varianzanalyse“ und „Mittelwertevergleiche“ bekannt. Viertens kann man statistische Zusammenhänge zwischen Gruppen und quantitativen Variablen zur Vorhersage der Gruppenzugehörigkeit nutzen, sofern der Wert der quantitativen Variablen bekannt ist. Untersuchungsmethoden mit diesem Ziel heißen „Diskriminanzanalyse“. Liegen nur qualitative Variablen vor, so kann man fünftens auftretende Häufigkeiten von Merkmalen mit der „Kontingenzanalyse“ auf statistische Abhängigkeit hin untersuchen.

Wir zeigen, daß alle fünf Verfahren zur Beschreibung von Zusammenhängen zwischen manifesten (d.h. beobachtbaren) Variablen auf der gleichen formalen Theorie basieren, die wir in Abschnitt 2.2 darstellen. Dabei gehen wir nicht auf Stichprobenprobleme wie Schätzverfahren und Hypothesenbeurteilung (z.B. durch Signifikanztests) ein. Diese sind Gegenstand der nachfolgenden Kapitel dieses Bandes.

## 2.2 Eine formale Theorie zur Beschreibung statistischer Zusammenhänge

### 2.2.1 Die Grundannahme

Die in diesem Abschnitt behandelte Klasse von Modellen zur Beschreibung des statistischen Zusammenhangs zwischen einer stochastischen Variablen  $y$  und mehreren stochastischen Variablen  $x_q$ ,  $q \in \{1, 2, \dots, Q\}$  basieren auf folgender linearen Gleichung

$$(2.1) \quad E(y | x_1, x_2, \dots, x_Q) = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \dots + \beta_Q \cdot x_Q$$

oder in Vektorschreibweise

$$E(y | x) = \beta_0 + \boldsymbol{\beta}' \mathbf{x},$$

wobei wir voraussetzen, daß alle beteiligten stochastischen Variablen auf dem gleichen Wahrscheinlichkeitsraum (s. Fußnoten 1 und 2) definiert sind und eine endliche Erwartung haben. Gleichung 2.1 besagt (siehe auch Anhang C), daß die bedingte Erwartung von  $y$  bei gegebenen  $x_1, x_2, \dots, x_Q$  eine Summe von  $\beta_0$  und einer Linearkombination der  $x_q$  ist. Die Variable  $y$  bezeichnet man häufig als abhängige Variable oder Kriteriumsvariable, die Variablen  $x_q$  als unabhängige Variablen oder Prädiktorvariablen<sup>3)</sup>.  $\beta_0$  ist eine Skalenkonstante, die angibt, welchen bedingten Erwartungswert die abhängige Variable  $y$  hat, wenn alle unabhängigen Variablen  $x_q$  den Wert Null haben. Die reellwertigen Koeffizienten  $\beta_q$ ,  $q \in \{1, 2, \dots, Q\}$  sind genau dann gleich Null, wenn kein durch Gleichung 2.1 beschriebener statistischer Zusammenhang zwischen  $y$  und  $x_q$  besteht. Die  $\beta_q$  wachsen in ihrem Absolutbetrag mit der „Enge“ des statistischen Zusammenhangs zwischen  $y$  und  $x_q$  bis zu einem Höchstbetrag, der von den Varianzen der beteiligten Variablen abhängig ist. Man beachte jedoch, daß diese vereinfachende Interpretation ihre Gültigkeit verliert, wenn „Suppressorvariablen“ (vgl. z.B. Darlington, 1968, S. 164, Fischer, 1974, S. 75, Conger & Jackson, 1972, Conger 1974) vorliegen.

### 2.2.2 Die Residualvariable

Statistische Zusammenhänge sind in der Regel nicht deterministisch. Zwischen der bedingten Erwartung  $E(y|x)$  und  $y$  selbst wird daher meist ein Unterschied bestehen. Diese Differenz, nämlich

---

<sup>3)</sup> Moosbrugger (1980) hat Gleichung 2.1 für nichtlineare multiple Variablenzusammenhänge mit Wechselwirkungen erweitert. Die bedingte Erwartung  $E(y | x) = \beta_0 + \boldsymbol{\beta}' \mathbf{z}$  ist dann eine Linearkombination der  $R \geq Q$  unabhängigen Variablen  $z_r, r \in \{1, \dots, R\}$ , welche beliebige Funktionen der bedingenden Variablen  $x_q$  sein können.

$$(2.2) \quad \varepsilon := y - E(y | \mathbf{x})$$

bezeichnen wir als Fehler- oder Residualvariable von  $y$ . Für die abhängige Variable  $y$  folgt dann aus den Gleichungen 2.1 und 2.2

$$(2.3) \quad y = \beta_0 + \boldsymbol{\beta}' \mathbf{x} + \varepsilon.$$

Für die Residualvariable von  $y$  gilt allgemein

$$(2.4) \quad E(\varepsilon | \mathbf{x}) = 0,$$

(siehe Anhang C, Regeln 4, 5 und 10), was bedeutet, daß die bedingten Erwartungswerte der Residualvariablen bei allen Ausprägungskombinationen der  $x_q$  gleich Null sind. Ohne diese Eigenschaft würde die stochastische Variable  $E(y | \mathbf{x})$  als Werte nicht die bedingten Erwartungswerte für  $y$  enthalten.

Eine weitere Eigenschaft der Residualvariablen ist, daß auch ihr unbedingter Erwartungswert

$$(2.5) \quad E(\varepsilon) = E[y - E(y | \mathbf{x})] = 0$$

gleich Null ist (siehe Anhang C, Regel 9), und daß die Kovarianzen zwischen allen unabhängigen Variablen  $x_q$  und der Residualvariablen  $\varepsilon$

$$(2.6) \quad C(\mathbf{x}, \varepsilon) = 0$$

gleich Null sind (siehe Anhang C, Regel 11). Man beachte, daß mit Gleichung 2.6 weder gefordert ist, daß die bedingte Fehlervarianz  $V(\varepsilon | \mathbf{x}=\mathbf{x})$  für alle Ausprägungskombinationen der  $x_q$  gleich ist, noch irgendeine spezielle Verteilungsannahme getroffen wird.

### 2.2.3 Kovarianzmodellgleichungen und Parameteridentifikation

Nachdem wir in Gleichung 2.1 ein statistisches Modell formuliert haben, stellt sich nun die Frage, ob die darin enthaltenen Parameter - in unserem Fall sind das  $\beta_0$  und die Komponenten von  $\boldsymbol{\beta}$  - eindeutig aus beobachtbaren Verteilungskennwerten der beteiligten stochastischen Variablen bestimmt (identifiziert) werden können. Im Fall eines Modells vom Typ der Gleichung 2.1 reicht die Betrachtung der Modellstruktur für den Kovarianzvektor  $C(x,y)$  sowie für die Varianzen  $V(y)$  und  $V(\varepsilon)$  aus.

Die Gleichungen 2.3 und 2.6 implizieren die folgenden Modellstrukturgleichungen

- a) für den Kovarianzvektor von  $x_q$  und  $y$

$$(2.7) \quad C(\mathbf{x}, y) = C[\mathbf{x}, (\beta_0 + \boldsymbol{\beta}'\mathbf{x} + \varepsilon)] = C(\mathbf{x}, \mathbf{x})\boldsymbol{\beta}$$

wobei  $C(\mathbf{x}, \mathbf{x})$  die Kovarianzmatrix der  $\mathbf{x}_q$  bezeichnet (siehe Anhang B, Regeln 1 und 2), und

b) für die Varianz von  $y$

$$(2.8) \quad V(y) = C[(\beta_0 + \boldsymbol{\beta}'\mathbf{x} + \varepsilon), (\beta_0 + \boldsymbol{\beta}'\mathbf{x} + \varepsilon)] = \boldsymbol{\beta}'C(\mathbf{x}, \mathbf{x})\boldsymbol{\beta} + V(\varepsilon)$$

(siehe Anhang B, Regeln 1 und 2).

Unter der zusätzlichen Annahme, daß  $C(\mathbf{x}, \mathbf{x})$  invertierbar ist, erhalten wir die Identifikationsgleichung für die Koeffizienten  $\beta_q$  aus Gleichung 2.7

$$(2.9) \quad \boldsymbol{\beta} = C(\mathbf{x}, \mathbf{x})^{-1} C(\mathbf{x}, y)$$

und jene für die Varianz der Residualvariablen  $\varepsilon$  aus den Gleichungen 2.8 und 2.9

$$(2.10) \quad V(\varepsilon) = V(y) - \boldsymbol{\beta}'C(\mathbf{x}, \mathbf{x})\boldsymbol{\beta}.$$

Wenn die Parameter  $\beta_q$  und die Residualvarianz  $V(\varepsilon)$  eindeutig bestimmt sind, lässt sich auch die unbekannte Skalenkonstante  $\beta_0$  leicht berechnen: Aus den Gleichungen 2.3 und 2.5 folgt nämlich nach Umformen (siehe Anhang A, Regel 2)

$$(2.11) \quad \beta_0 = E(y) - \boldsymbol{\beta}'E(\mathbf{x})$$

als Differenz des Erwartungswertes von  $y$  und den mit den Koeffizienten  $\beta_q$  gewichteten Erwartungswerten der  $\mathbf{x}_q$ .

Lassen sich wie hier alle unbekannten Modellparameter aus den Erwartungswerten, Varianzen und Kovarianzen der beobachtbaren Variablen eindeutig bestimmen, so bezeichnet man das Modell als identifiziert. Man beachte, daß es sich bei den Gleichungen 2.9 bis 2.11 um Identifikationsgleichungen handelt. Schätzgleichungen, in denen analoge Stichprobenkennwerte verwendet werden, finden sich z.B. in Moosbrugger (1978, S. 56 und 61). über andere Schätzmethoden informieren z.B. Späth (1973, 1974), Harter (1974a,b, 1975a,b,c), Hoerl & Kennard (1970) oder Humak (1977).

Wenn  $C(\mathbf{x}, \mathbf{x})$  nicht invertierbar ist, kann eine allgemeine Lösung für  $\boldsymbol{\beta}$  immer durch

$$(2.12) \quad \boldsymbol{\beta} = C(\mathbf{x}, \mathbf{x})^- C(\mathbf{x}, y)$$

gefunden werden, wobei  $C(\mathbf{x}, \mathbf{x})^-$  die verallgemeinerte Inverse (vgl. z.B. Rao & Mitra, 1971) der Kovarianzmatrix  $C(\mathbf{x}, \mathbf{x})$  ist. Eine solche Lösung wird im

allgemeinen jedoch erst durch zusätzliche Nebenbedingungen eindeutig (siehe z.B. Moosbrugger & Steyer (1982) in diesem Band).

#### 2.2.4 Determinierte Varianz, multiple Korrelation und Determinationskoeffizient

Wir bezeichnen mit  $V [E(y | \mathbf{x})]$  die durch  $\mathbf{x}$  determinierte Varianz von  $y$ . Unter Verwendung von Gleichung 2.1 erhalten wir die Beziehung

$$(2.13) \quad V [E(y | \mathbf{x})] = C(\boldsymbol{\beta}' \mathbf{x}, \boldsymbol{\beta}' \mathbf{x}) = \boldsymbol{\beta}' C(\mathbf{x}, \mathbf{x}) \boldsymbol{\beta}.$$

Setzen wir Gleichung 2.13 in Gleichung 2.10 ein, so sehen wir, daß die determinierte Varianz die Differenz von Gesamtvarianz und Residualvarianz ist:

$$(2.14) \quad V [E(y | \mathbf{x})] = V(y) - V(\varepsilon).$$

Dividieren wir Gleichung 2.14 durch  $V(y)$ , so erhalten wir unter Verwendung von Gleichung 2.13 und unter der Voraussetzung  $V(y) \neq 0$  den durch  $\mathbf{x}$  determinierten Varianzanteil von  $y$ :

$$(2.15) \quad \frac{V [E(y | \mathbf{x})]}{V(y)} = \frac{V(y) - V(\varepsilon)}{V(y)} = \frac{\boldsymbol{\beta}' C(\mathbf{x}, \mathbf{x}) \boldsymbol{\beta}}{V(y)} =: R^2(y, \mathbf{x}).$$

Dieser Kennwert heißt multipler Determinationskoeffizient, die positive Wurzel daraus multiple Korrelation (vgl. auch Moosbrugger, 1978, S. 65). Die Werte von  $R^2(y, \mathbf{x})$  liegen zwischen Null und Eins, wobei  $R^2(y, \mathbf{x})$  den Wert Null annimmt, wenn  $\boldsymbol{\beta}$  bzw.  $C(\mathbf{x}, y)$  der Nullvektor ist.

#### 2.2.5 Multivariate Verallgemeinerung

Liegen  $P$  abhängige Variablen vor, so können wir die Grundannahme 2.1 durch deren multivariate Verallgemeinerung

$$(2.16) \quad E(\mathbf{y} | \mathbf{x}) = \boldsymbol{\beta}_0 + \mathbf{B}' \mathbf{x}$$

ersetzen. Entsprechend dem Vorgehen beim univariaten ( $P = 1$ ) Fall erhalten wir hier die Modellgleichung für die Kovarianzmatrizen

$$(2.17) \quad C(\mathbf{x}, \mathbf{y}) = C(\mathbf{x}, \mathbf{x}) \mathbf{B}$$

und unter der Voraussetzung der Invertierbarkeit von  $C(\mathbf{x}, \mathbf{x})$ , die Identifikationsgleichungen

$$(2.18) \quad \mathbf{B} = C(\mathbf{x}, \mathbf{x})^{-1} C(\mathbf{x}, \mathbf{y})$$

und

$$(2.19) \quad \beta_0 = E(y) - \mathbf{B}'E(x).$$

Andernfalls gilt wieder die allgemeine Lösung

$$(2.20) \quad \mathbf{B} = C(x,x)^{-1} C(x,y),$$

die erst durch zusätzlich Nebenbedingungen eindeutig wird.

Als multivariate Verallgemeinerung des multiplen Determinationskoeffizienten kann der multivariate Determinationskoeffizient

$$(2.21) \quad R^2(y,x) = \frac{\text{Spur} \{ C[E(y|x), E(y|x)] C(y,y)^{-1} \}}{P}$$

angesehen werden (vgl. auch Cramer & Nicewander, 1979), dessen Werte zwischen Null und Eins liegen. Dabei ist

$$P = \text{Spur} [C(y,y) C(y,y)^{-1}] = \text{Spur I}$$

gleich der Anzahl der abhängigen Variablen. Näheres siehe Moosbrugger & Steyer (1982) in diesem Band.

## 2.2.6 Zusammenfassende Bemerkungen

Die in diesem Abschnitt dargestellte Theorie beruht auf einer einzigen Annahme, nämlich der Gleichung 2.1. Daraus läßt sich die Identifikationsgleichung für den dabei auftretenden Parametervektor  $\beta$  ableiten, falls  $C(x,x)$  invertierbar ist. Andernfalls müssen zusätzliche Nebenbedingungen über die Parameter in  $\beta$  formuliert werden, um zu einer eindeutigen Lösung für  $\beta$  aus den Kovarianzmatrizen  $C(x,x)$  und  $C(x,y)$  zu gelangen. Analoges gilt für den multivariaten Fall. Mit dem multiplen - bzw. multivariaten Determinationskoeffizienten kann angegeben werden, wie eng der statistische Zusammenhang zwischen  $y$  (bzw.  $y$ ) und  $x$  ist. In den nachfolgenden Abschnitten wollen wir verdeutlichen, daß die beschriebenen Konzepte in verschiedensten Situationen der psychologischen Forschung ihre Anwendung finden.

## 2.3 Anwendungen der formalen Theorie.

In den folgenden Unterkapiteln wollen wir zeigen, um welche Spezialfälle der in Kapitel 2.2 dargestellten allgemeinen Theorie es sich bei der Regressions-, Korrelations-, Varianz-, Diskriminanz- und Kontingenzanalyse handelt. Dabei werden auch die Zusammenhänge zwischen den genannten Analyseverfahren deutlich.

### 2.3.1 Regressions- und Korrelationsanalyse

Die Begriffe „Regression“ und „Korrelation“ entstanden im letzten Jahrhundert. Explizit wurden „Regression“ und „Co-Relation“ von Galton zumindest seit 1889 benutzt (Galton 1889). Galtons erste, aber noch nicht als solche bezeichnete „Regressionslinie“ vom Jahre 1877 zeigt die Abhängigkeit des Durchmessers von Blatterbsensamen von jenen der zugehörigen Mutterpflanzen. Mit der Größe des Samendurchmessers der Mutterpflanzen steigt jener der Tochterpflanzen an, bei letzteren jedoch in geringerem Ausmaß. Diese Beobachtung, daß die Durchmesser von Blatterbsensamen der Tochterpflanzen dazu tendieren, zur mittleren Größe des Samendurchmessers der Vorfahren zurückzukehren, bezeichnete Galton 1889 als „Regression“. Regressions- und Korrelationsmethoden wurden dann von Pearson (1896) sowie Yule (1897) weiter entwickelt. Doch bereits lange vor Galton wurden „Regressionsanalysen“ durchgeführt, freilich ohne daß dabei dieser Begriff Verwendung fand, so z.B. auf geophysikalischem Gebiet von Adrain (1818) oder von den Wiener Astronomen Kreil (1831) und Littrow (1818, 1833), welche bereits genaue Berechnungsformeln der „Regressionskoeffizienten“ nach der Minimum-Quadrat-Methode angaben. (Ausführlichere geschichtliche Angaben machen z.B. Weiling, 1972, 1973 und David, 1978).

In der Regressionsanalyse handelt es sich bei  $y$  und den  $x_q$  in der Regel um quantitative stochastische Variablen. Sind z.B. die Meßwerte einer  $V_p$  in den Variablen  $x_q$  gegeben und sind die Koeffizienten  $\beta_q$  in der Gleichung 2.1 bekannt, so läßt sich der Wert  $y$  der Variablen  $y$  mit einem bestimmten Fehler  $\varepsilon$  voraussagen. Jeder für einen festen Vektor  $x$  vorhergesagte Erwartungswert  $E(y | \mathbf{x}=\mathbf{x})$  ist eine Realisation der bedingten Erwartung  $E(y | x)$ , welche im Fall kontinuierlicher Variablen und linearer Beziehungen zwischen  $y$  und den  $x_q$  eine Ebene bzw. eine Hyperebene ist und Regressionsebene heißt. Für zwei unabhängige Variablen beschreibt  $E(y|x_1, x_2)$  eine Ebene, welche zur Prädiktion von  $y$  aus  $x_1$  und  $x_2$  benutzt werden kann (siehe Abbildung 2.1). Dabei ist  $\beta_0$  der Ordinatenabschnitt, also die Höhe, in der die Regressionsebene die  $y$ -Achse schneidet,  $\beta_1$  die Steigung der Ebene entlang der  $x_1$ -Achse und  $\beta_2$  die Steigung entlang der  $x_2$ -Achse.

Das Ausmaß der Vorhersagbarkeit von  $y$  hängt von der „Enge“ des linearen statistischen Zusammenhangs ab, die wir mit dem multiplen Determinationskoeffizienten angeben können (siehe Gleichung 2.15).

Im Fall eines multiplen Regressionsmodells mit zwei Prädiktoren und Interaktion,

$$(2.22) \quad E(y | x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$$

ist die bedingte Erwartung  $E(y | x_1, x_2)$  eine gekrümmte Fläche (siehe Abbildung 2.2), deren Steigung entlang der  $x_1$ -Achse auf der Stufe  $x_2=0$  der Parame-

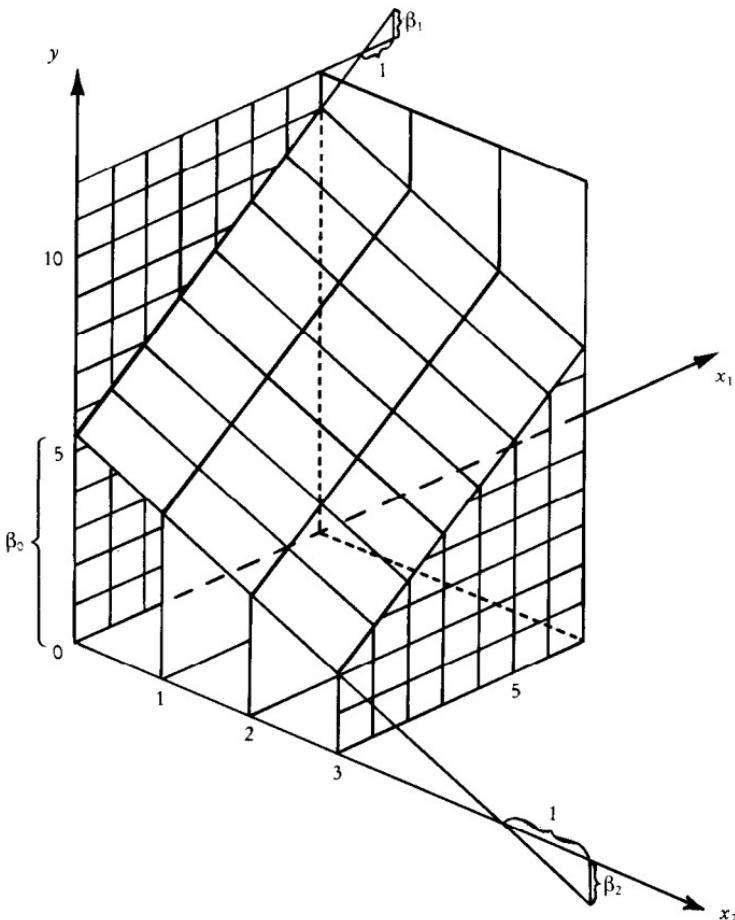


Abb. 2.1: Regressionsmodell mit zwei Prädiktoren

$$E(y | x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

mit

$$\beta_0 = 5.5$$

$$\beta_1 = 0.8$$

$$\beta_2 = -1.1$$

ter  $\beta_1$  angibt und entlang der  $x_2$ -Achse auf der Stufe  $x_1 = 0$  der Parameter  $\beta_2$ . Der Parameter  $\beta_3$  gibt jenen Anteil der bedingten Erwartung an, welcher auf bestimmte Kombinationen von Merkmalsausprägungen in den Variablen  $x_1$  und  $x_2$  zurückgeht: Von Stufe zu Stufe der Variablen  $x_2$  (bzw.  $x_1$ ) ist die Steigung der Variablen  $x_1$  (bzw.  $x_2$ ) um  $\beta_3$  verschieden (moderiert). Somit ist  $\beta_3$  genau der Unterschied zwischen den Steigungen der Regressionsgeraden von  $y$  auf  $x_1$  (bzw.  $x_2$ ) in jeweils zwei um eine Einheit entfernt liegenden Ausprägungen der Variablen  $x_2$  (bzw.  $x_1$ ). Daher wird  $x_2$  (bzw.  $x_1$ ) als Moderatorvariable

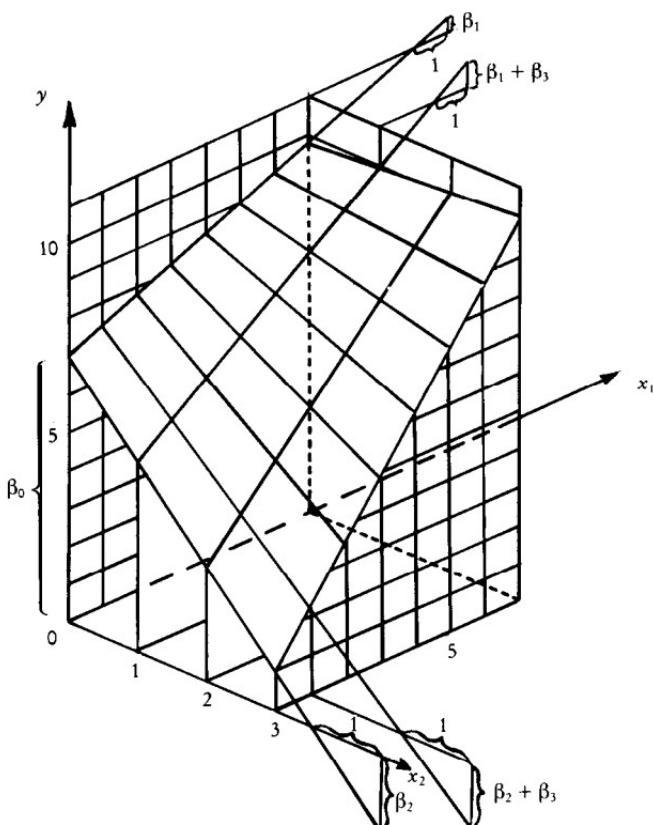


Abb. 2.2: Regressionsmodell mit zwei Prädiktoren und Interaktion  
 $E(y | x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$   
mit

$$\begin{aligned}\beta_0 &= 7.0 \\ \beta_1 &= 0.4 \\ \beta_2 &= -2.0 \\ \beta_3 &= 0.3\end{aligned}$$

bezeichnet, wenn  $\beta_3 \neq 0$  gilt (vgl. Bartussek, 1970, S. 11, Moosbrugger, 1981 a, S. 220).

Haben wir eine Stichprobe von Werten der Variablen  $y$  und  $x_q$  vorliegen, so können wir für multiple Regressionsmodelle unter der Normalverteilungsannahme der Residualvariablen Hypothesen prüfen, indem wir berechnen, wie wahrscheinlich es ist, daß alle oder einzelne  $\beta_q$  in der Population gleich Null sind (siehe z.B. Moosbrugger, 1978, S. 72-77), oder daß sie bzw. Linear-kombinationen aus ihnen andere hypothetische feste Werte annehmen (siehe z.B. Timm, 1975, S. 177ff. oder Moosbrugger & Steyer, 1982, in diesem Band).

Weiterführende Literatur: Die Regressionsanalyse gehört zu den Themen, die von den meisten Lehrbüchern der angewandten Statistik behandelt werden. Ausführlichere Einführungen in die Regressionsanalyse geben Moosbrugger (1978), sowie Gaenssen & Schubö (1973), deren Darstellung Matrixalgebra vermeidet. Des weiteren ist auch Kerlinger & Pedhazur (1973) zur Einführung geeignet, außerdem Edwards (1976) sowie Draper & Smith (1966). Auf die Bedeutung von komplexen Regressionsmodellen mit Moderatorvariablen für die differentielle Validität psychologischer Tests geht z.B. Moosbrugger (1981 a) ein.

### 2.3.2 Varianzanalyse

Die Varianzanalyse wurde von Sir R. A. Fisher (1925) für die Auswertung von landwirtschaftswissenschaftlichen Experimenten entwickelt, bei denen es z.B. um die Effekte von Düngemitteln auf den Ertrag ging. Nach dem zweiten Weltkrieg hat dieses Verfahren jedoch auch in der Psychologie verstärkt Anwendung gefunden, wobei in der Regel die Wirkungen experimenteller Manipulationen überprüft werden. Heute zählt sie zu den Standardverfahren der psychologischen Forschungsmethoden. Zur Geschichte der Varianzanalyse siehe Weiling (1973).

Während bei varianzanalytischen Modellen mit Zufallsparametern (random-model, vgl. z.B. Hays, 1973, Kap. 13) die  $\beta_q$  in Gleichung 2.1 stochastische Variablen wären, können Varianz- und kovarianzanalytische Modelle mit festen Parametern (fixed-model, vgl. z.B. Hays, 1973, Kap. 12) als Spezialfalle der Gleichung 2.1 angesehen werden. Wir beschränken uns daher auf Modelle mit festen Parametern. Bei  $y$  handelt es sich dann wieder um eine quantitative stochastische Variable, bei den  $x_q$  jedoch in der Varianzanalyse ausschließlich um qualitative Indikatorvariablen (synonym auch: *Kodier-* oder *Dummyvariablen*), deren Werte die Gruppenzugehörigkeit anzeigen, und in der *Kovarianzanalyse* (siehe z.B. Moosbrugger, 1978, Kap. 15) um qualitative und quantitative Variablen.

In der einfaktoriellen Varianzanalyse mit zwei Stufen beispielsweise liegen als Werte der abhängigen Variablen  $y$  eine Reihe quantitativer Meßwerte vor, sowie die Information, welcher von zwei Gruppen ein Wert  $y$  jeweils angehört. Von Interesse ist dann der Mittelwerteunterschied von  $y$  in beiden Gruppen (z.B. Behandlungs- und Kontrollgruppe). Kodieren wir die Zugehörigkeit zur ersten Gruppe mit  $x = 1$  und die zur zweiten Gruppe mit  $x = -1$ , so ergibt sich das Datenschema der Abbildung 2.3.<sup>4)</sup>

<sup>4)</sup> Man beachte dabei, daß eine Kodierung der Gruppenzugehörigkeit mit 1 und 0 ebenfalls zu gut interpretierbaren Parametern führen würde (vgl. auch Rochel, 1981). Indikationsfragen verschiedener Kodierungsverfahren wurden von Schermelleh (1979) untersucht.

G r u p p e		$y$	$x$
1	$y_{11}$		1
	$y_{12}$		1
	.		.
	.		.
	.		.
	$y_{1N_1}$		1
2	$y_{21}$		-1
	$y_{22}$		-1
	.		.
	.		.
	.		.
	$y_{2N_2}$		-1

Abb. 2.3: Datenschema eines einfaktoriellen varianzanalytischen Modells mit zwei Stufen,  $\mathbf{E}(\mathbf{y} | \mathbf{x}) = \beta_0 + \boldsymbol{\beta} \mathbf{x}$ . Die abhängige Variable ist  $y$ , die an  $N_1$  bzw.  $N_2$  Beobachtungseinheiten in den beiden Gruppen erhoben wird. Die Indikatorvariable  $x$  enthält die Information über die Gruppenzugehörigkeit der Beobachtungseinheiten.

In der *zweifaktoriellen* Varianzanalyse mit je zwei Stufen (siehe Abbildung 2.4) beispielsweise liegen die Werte der abhängigen Variablen  $y$  vor sowie die Information, aus welcher von vier Gruppen ein Wert  $y$  jeweils stammt. Von Interesse sind lineare oder nichtlineare statistische Zusammenhänge zwischen den  $x_q$  und  $y$ , welche sich hier als Unterschiede der Erwartungswerte von  $y$  in den vier Gruppen zeigen.

		Faktor A	
		Stufe 1	Stufe 2
Faktor B	Stufe 1	1. Gruppe	3. Gruppe
	Stufe 2	2. Gruppe	4. Gruppe

Abb. 2.4: Zweifaktorieller vollständig gekreuzter Versuchsplan mit 2 Stufen pro Faktor

Wir kodieren die zwei Stufen des Faktors A (siehe Abbildung 2.4) durch die Kodiervariable  $x_1$  mit den Werten  $x_1 = 1$  für die Stufe 1 und  $x_1 = -1$  für Stufe

2, und die zwei Stufen des Faktors B durch die Kodiervariable  $x_2$  mit den Werten  $x_2 = 1$  für die Stufe 1 und  $x_2 = -1$  für Stufe 2. Liegen mehr als zwei Stufen pro Faktor vor, benötigt man auch mehr als eine Indikatorvariable pro Faktor. Man beachte, daß die Art der Kodierung zwar die Interpretation der Parameter bestimmt, aber keinen Einfluß auf die Hypothesenbeurteilung hat (siehe z.B. Moosbrugger, 1978, Moosbrugger & Steyer (1982) in diesem Band, Rochel, 1981, Steyer, 1979 oder Wottawa, 1974).

Nach dem in Abbildung 2.4 schematisierten kreuzfaktoriellen Versuchsplan ergibt sich für die vier Subpopulationen das in Abbildung 2.5 dargestellte Datenschema:

Gruppe	$y$	$x_1$	$x_2$	$x_1 \cdot x_2$
1	$y_{11}$	1	1	1
	$y_{12}$	1	1	1
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$y_{1N_1}$	1	1	1
2	$y_{21}$	1	-1	-1
	$y_{22}$	1	-1	-1
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$y_{2N_2}$	1	-1	-1
3	$y_{31}$	-1	1	-1
	$y_{32}$	-1	1	-1
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$y_{3N_3}$	-1	1	-1
4	$y_{41}$	-1	-1	1
	$y_{42}$	-1	-1	1
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$y_{4N_4}$	-1	-1	1

Abb. 2.5: Varianzanalytisches Datenschema zum Versuchsplan in Abbildung 2.4.

In der ersten Spalte ist die abhängige Variable  $y$  notiert, die an  $N_g$  Beobachtungseinheiten in der Gruppe  $g$ ,  $g \in \{1,2,3,4\}$ , erhoben wird. Die Spalten 2 und 3 zeigen die Indikatorvariablen  $x_1$  und  $x_2$  für die Stufen der Faktoren A und B. Die Variable  $x_1 \cdot x_2$  in Spalte 4 ist ein Indikator für bestimmte Faktorstufenkombinationen. Sie dient der Erfassung von Wechselwirkungen (Interaktionen, s.u.). Wird keine Interaktion angenommen, so entfällt die Spalte  $x_1 \cdot x_2$ .

Wie in der Regressionsanalyse liegen uns mit den Datenschemata der Abbildungen 2.3 und 2.5 Werte von stochastischen Variablen vor, deren statistischer Zusammenhang mit Gleichung 2.1 beschrieben werden kann.

Wir untersuchen im folgenden am zweifaktoriellen Design, welche Interpretation den Parametern  $\beta_q$  in diesem Kontext zukommt, zunächst ohne, danach mit Annahme einer Interaktion (Wechselwirkung). Ein statistischer Zusammenhang zwischen  $y$  und den  $x_q$  ohne Interaktion liegt dann vor, wenn Mittelwerteunterschiede zwischen den Gruppen derart bestehen, daß die Differenz der Erwartungswerte zwischen den Stufen des Faktors A auf jeder der Stufen des Faktors B gleich groß ist und umgekehrt.

In welcher Beziehung die Parameter  $\beta_0$  und  $\beta_q$  der Gleichung

$$(2.23) \quad E(y | x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

mit den Erwartungswerten von  $y$  in den Gruppen stehen, sehen wir, wenn wir die bedingten Erwartungswerte von  $y$  für die vier Gruppen bilden:

$$(2.24) \quad E(y_1) := E(y | x_1 = 1, x_2 = 1) = \beta_0 + \beta_1 + \beta_2$$

$$(2.25) \quad E(y_2) := E(y | x_1 = 1, x_2 = -1) = \beta_0 + \beta_1 - \beta_2$$

$$(2.26) \quad E(y_3) := E(y | x_1 = -1, x_2 = 1) = \beta_0 - \beta_1 + \beta_2$$

$$(2.27) \quad E(y_4) := E(y | x_1 = -1, x_2 = -1) = \beta_0 - \beta_1 - \beta_2.$$

Gewöhnlich werden die Gruppen als gleich groß angenommen, was bedeutet, daß die Wahrscheinlichkeiten für die Faktorstufen  $P(x_1 = 1) = P(x_1 = -1) = 1/2$  und  $P(x_2 = 1) = P(x_2 = -1) = 1/2$  gilt. Dies stellt sicher, daß die unabhängigen Variablen  $x_1$  und  $x_2$  zueinander orthogonal sind, d.h. nicht miteinander korrelieren, so daß deren Einflüsse auf die abhängige Variable  $y$  voneinander unabhängig geschätzt werden können. Ist dies nicht der Fall, handelt es sich um eine nonorthogonale Varianzanalyse (siehe z.B. Steyer, 1979).

Unter den obigen Annahmen der Gleichwahrscheinlichkeit für die Faktorausprägungen finden wir Null als Erwartungswert der Indikatorvariablen:

$$(2.28) \quad E(x_1) = P(x_1 = 1) \cdot 1 + P(x_1 = -1) \cdot -1 = 0$$

und

$$(2.29) \quad E(x_2) = P(x_2 = 1) \cdot 1 + P(x_2 = -1) \cdot -1 = 0.$$

Bilden wir den Erwartungswert von Gleichung 2.23, so erhalten wir dann (vgl. Anhang C, Regel 1)

$$(2.30) \quad \beta_0 = E(y) - \beta_1 E(x_1) - \beta_2 E(x_2) = E(y).$$

Demnach ist  $\beta_0$  unter der genannten Voraussetzung gleich dem Erwartungswert von  $y$ . Einsetzen von Gleichung 2.30 in die Gleichungen 2.24 bis 2.27 führt zu den Lösungen für

$$(2.31) \quad \beta_1 = \frac{E(y_1) + E(y_2)}{2} - E(y) = E(y) - \frac{E(y_3) + E(y_4)}{2}$$

und

$$(2.32) \quad \beta_2 = \frac{E(y_1) + E(y_3)}{2} - E(y) = E(y) - \frac{E(y_2) + E(y_4)}{2}.$$

Wir sehen, daß  $\beta_1$  den „Haupteffekt von Faktor A“, (vgl. z.B. Hays, 1973, S. 494), oder, bezogen auf Abbildung 2.4, den „Spalteneffekt“ darstellt, wohingegen  $\beta_2$  der „Haupteffekt von Faktor B“, oder bezogen auf Abbildung 2.4, der „Zeileneffekt“ ist<sup>5</sup>). Abbildung 2.6 veranschaulicht die Beziehungen zwischen den Parametern und den Erwartungswerten in den vier Gruppen:

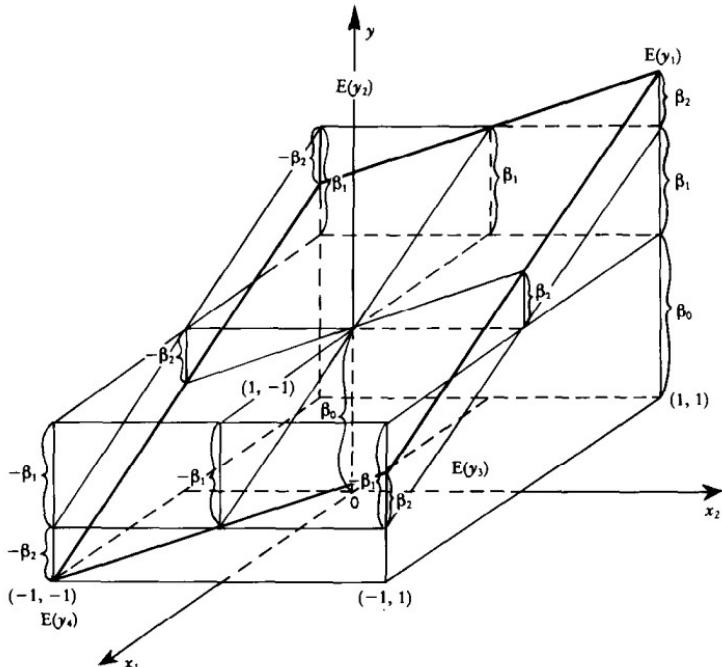


Abb. 2.6: Veranschaulichung der Erwartungswerte von  $y$  in den 4 Gruppen eines zweifaktoriellen varianzanalytischen Modells gemäß den Gleichungen 2.24 bis 2.27 mit  $\beta_0 = 3$ ,  $\beta_1 = 2$  und  $\beta_2 = 1$ .

<sup>5</sup>) Die weitverbreitete Bezeichnung „Effekt“ impliziert eine kausale Interpretation, welche man strenggenommen nur im Kontext eines kausalen Modells (siehe Steyer, 1982, in diesem Band) vornehmen sollte.

Ist der statistische Zusammenhang zwischen den bedingenden Variablen  $x_q$  und der abhängigen Variablen  $y$  nichtlinear in den bedingenden Variablen<sup>6)</sup>, so moderieren (verändern) die Stufen des Faktors B den Effekt des Faktors A oder umgekehrt (vgl. auch Kap. 2.3.1). Einen solchen Moderatoreffekt, der in der Varianzanalyse Wechselwirkungs- oder Interaktionseffekt heißt, können wir mit einer weiteren Indikatorvariablen  $x_1 \cdot x_2$  für die Faktorstufenkombinationen und dem Parameter  $\beta_3$  erfassen. Die Spalte 4 in Abbildung 2.5 (s.o.) zeigt die entsprechenden Werte von  $x_1 \cdot x_2$ . Ein Haupteffekt stellt dann den durchschnittlichen Effekt eines Faktors gemittelt über die Stufen des anderen Faktors dar. Für die bedingte Erwartung von  $y$  unter  $x$  gilt die folgende Gleichung

$$(2.33) \quad E(y | x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 \cdot x_2.$$

Zur Untersuchung, in welcher Beziehung in diesem Fall die Parameter  $\beta_0$  und  $\beta_3$  der Gleichung 2.33 mit den Gruppenerwartungswerten stehen, bilden wir wieder die bedingten Erwartungswerte von  $y$  für die vier Gruppen:

$$(2.34) \quad E(y_1) := E(y | x_1 = 1, x_2 = 1) = \beta_0 + \beta_1 + \beta_2 + \beta_3$$

$$(2.35) \quad E(y_2) := E(y | x_1 = 1, x_2 = -1) = \beta_0 + \beta_1 - \beta_2 - \beta_3$$

$$(2.36) \quad E(y_3) := E(y | x_1 = -1, x_2 = 1) = \beta_0 - \beta_1 + \beta_2 - \beta_3$$

$$(2.37) \quad E(y_4) := E(y | x_1 = -1, x_2 = -1) = \beta_0 - \beta_1 - \beta_2 + \beta_3.$$

Bei gleichgroßen Subpopulationen gelten außer den Gleichungen 2.28 und 2.29 auch  $P(x_1 \cdot x_2 = 1) = P(x_1 \cdot x_2 = -1) = 1/2$  und  $E(x_1 \cdot x_2) = 0$ , was wegen  $E(x_1) = E(x_2) = 0$  wiederum (siehe Gleichung 2.30) zu

$$(2.38) \quad \beta_0 = E(y)$$

führt.

Einsetzen von Gleichung 2.38 in die Gleichungen 2.34 bis 2.37 ergibt unveränderte Ausdrücke für  $\beta_1$  und  $\beta_2$  (siehe Gleichungen 2.31 und 2.32) und zusätzlich eine Gleichung für  $\beta_3$ :

$$(2.39) \quad \beta_3 = \frac{E(y_1) + E(y_4)}{2} - E(y) = E(y) - \frac{E(y_2) + E(y_3)}{2}.$$

Der Interaktionsparameter, nämlich der Koeffizient  $\beta_3$  der Variablen  $x_1 \cdot x_2$  ist genau dann Null, wenn der Erwartungswert von  $y$  gleich dem arithmetischen

<sup>6)</sup> Zur Klassifikation von linearen vs. nichtlinearen Modellen hinsichtlich der Parameter  $\beta_q$  bzw. linearen vs. nichtlinearen Modellen hinsichtlich den bedingenden Variablen  $x_q$  siehe Moosbrugger, 1980, S. 5-12.

Mittel der Erwartungswerte in den Gruppen 1 und 4 bzw. 2 und 3 ist. Dies ist der Fall, wenn sich die Erwartungswerte von  $y$  in den vier Gruppen allein aus dem Gesamterwartungswert  $E(y) = \beta_0$  und den Parametern  $\beta_1$  und  $\beta_2$  darstellen lassen. Die Größe von  $\beta_3$  wächst mit der Nichtlinearität, nämlich mit der Abweichung des Durchschnitts von je zwei, bezogen auf die Abbildung 2.4 „diagonalen“, Gruppenerwartungswerten vom Gesamterwartungswert. Der Veranschaulichung dient Abbildung 2.7:

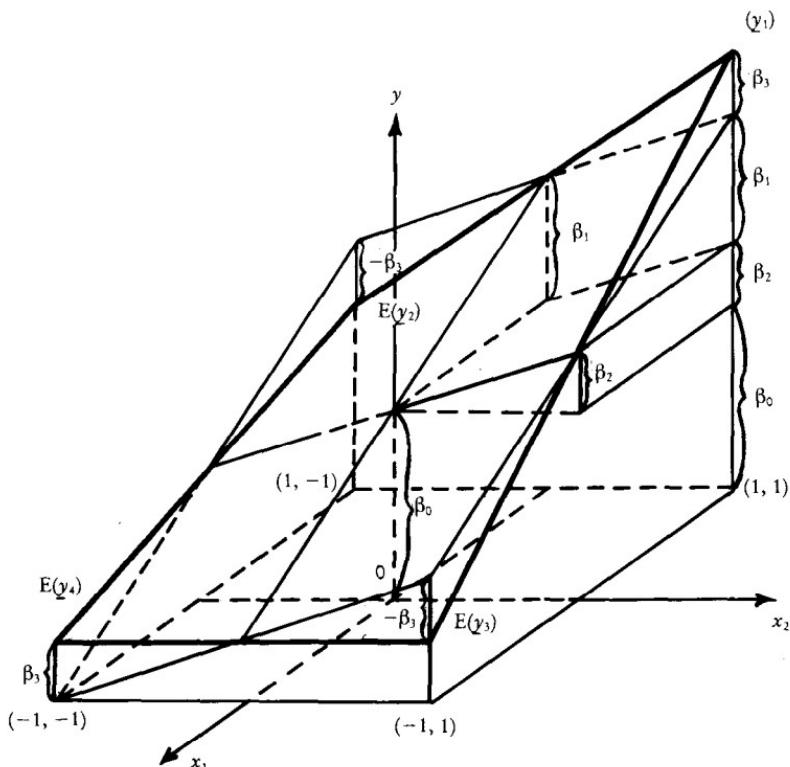


Abb. 2.7: Veranschaulichung der Erwartungswerte von  $y$  in den 4 Subpopulationen eines zweifaktoriellen varianzanalytischen Modells mit Wechselwirkung gemäß den Gleichungen 2.34 bis 2.37 mit  $\beta_0 = 3$ ,  $\beta_1 = 2$ ,  $\beta_2 = 1$  und  $\beta_3 = 1$  im Unterschied zu  $E(y|x_1, x_2)$  in Abbildung 2.6. Der Haupteffekt  $\beta_1 = 2$  gibt hier nur den durchschnittlichen Effekt von Faktor A an, welcher durch die Stufen von Faktor B moderiert wird. Die Steigung (der Effekt) auf der Stufe  $B_1$  beträgt  $(\beta_1 + \beta_3) = 3$ , auf der Stufe  $B_2$  hingegen  $(\beta_1 - \beta_3) = 1$ . Die analoge Aussage gilt für Faktor B: auf der Stufe  $A_1$  beträgt der Effekt  $(\beta_2 + \beta_3) = 2$ , auf der Stufe  $A_2$  hingegen  $(\beta_2 - \beta_3) = 0$ ; den Durchschnitt zeigt der Haupteffekt  $\beta_2 = 1$  an.

Ebenso wie bei der Regressionsanalyse können wir in der Varianzanalyse unter der Normalverteilungsannahme Hypothesen prüfen, indem wir berechnen, wie wahrscheinlich es ist, daß alle oder einzelne  $\beta_q$  in der Population gleich Null sind (siehe z.B. Moosbrugger, 1978, S. 72-77), oder daß Linearkombinationen der  $\beta_q$  hypothetische feste Werte annehmen (siehe z.B. Timm, 1975, S. 177ff. oder Moosbrugger & Steyer, 1982, in diesem Band), falls wir eine Stichprobe von Werten der Variablen  $y$  und  $x_q$  vorliegen haben.

Weiterführende Literatur: Auch die Varianzanalyse gehört zu den Standardthemen der Lehrbücher für angewandte Statistik. Außer den bereits unter der Regressionsanalyse (s.o.) genannten Büchern seien Searle (1971), Timm (1975), Bock (1975) und Finn (1974) aufgeführt. Als ausgesprochenes Standardwerk gilt Winer (1971<sup>2</sup>), der für sehr viele Designs Rechenformeln angibt. Als weitere deutschsprachige Bücher seien Bandemer et. al. (1977), Bortz (1977) und Rasch (1976) genannt. Eine Einführung in die multivariate Varianzanalyse geben Moosbrugger & Steyer (1982) in diesem Band.

### 2.3.3 Diskriminanzanalyse

Auch die Diskriminanzanalyse wurde von R. A. Fisher (1936) entwickelt. Sein Anwendungsbeispiel war die Zuordnung von Irispflanzen zu einer von zwei Arten auf der Basis der Länge und Breite der Kelch- und Blütenblätter (Anderson, 1978, S. 628). Die typische Fragestellung der Diskriminanzanalyse in der Psychologie ist die der Klassifikation, nämlich, zu welcher von bestimmten (z.B. Krankheits-)Gruppen eine Person gehört, von der bestimmte Informationen  $x$  (z.B. ein oder mehrere Testwerte) vorliegen.

Formal gesehen ist die Diskriminanzanalyse eigentlich eine Umkehrung der Varianzanalyse. Ist es nach einer Varianzanalyse mit bedeutsamen Mittelwertunterschieden möglich, von der Kenntnis der Gruppenzugehörigkeit der qualitativen Variablen auf die durchschnittliche Ausprägung der quantitativen Variablen zu schließen, so ermöglicht in Umkehrung dieses Vorgangs die sogenannte Diskriminanzfunktion einen Schluß, in welche der Gruppen eine bestimmte Wertekombination der quantitativen Variablen am ehesten klassifiziert werden kann.

Die Grundgedanken lassen sich am leichtesten für den Fall zweier Gruppen darstellen. Hierbei ist zu beachten, daß wir, um die formale Verträglichkeit mit der Gleichung 2.1 herbeizuführen, gegenüber der Varianzanalyse die Bedeutung von  $x$  und  $y$  vertauschen. Nun sei  $y$  eine Indikatorvariable, wobei  $y = 1$  Zugehörigkeit zur Gruppe 1 und  $y = 0$  Zugehörigkeit zur Gruppe 2 bedeuten soll. Die quantitativen Informationen seien in den kontinuierlichen Variablen  $x_q$  gesammelt. Da wir in unserer bisherigen Erörterung nur eine abhängige Variable gegenüber mehreren unabhängigen zugelassen haben, ist mit einer Kodiervariablen  $y$  eine Unterscheidung von nur zwei Gruppen mög-

lich. Mehrere Gruppen erfordern ein multivariates Modell (siehe z.B. Bock, 1975, S. 403), um mit mehreren Indikatorvariablen  $y_p$  mehr als zwei Gruppen kodieren zu können.

In der Diskriminanzanalyse heißt die bedingte Erwartung  $E(y|x)$  auch Diskriminanzfunktion. Wird diese wiederum als linear angenommen, so gilt weiterhin Gleichung 2.1, doch mit der oben erwähnten Vertauschung von  $x$  und  $y$  geht auch eine Veränderung der Bedeutung der Parameter  $\beta_q$  einher, welche wir kurz untersuchen wollen:

Die Interpretation der Skalenkonstanten  $\beta_0$  wird erleichtert, wenn wir die Erwartungswerte aller kontinuierlichen Variablen  $x_q$ ,  $q \in \{1,2,\dots,Q\}$ , gleich Null setzen. Dann ist

$$(2.40) \quad E(y) = E [E(y|x)] = \beta_0 + \beta_1 E(x_1) + \dots + \beta_Q E(x_Q) = \beta_0 \cdot$$

Da die Gruppenvariable  $y$  nur die Werte  $y = 1$  und  $y = 0$  annehmen kann, gilt für ihren Erwartungswert und folglich für  $\beta_0$

$$(2.41) \quad \beta_0 = E(y) = P(y=1) \cdot 1 + P(y=0) \cdot 0 = P(y=1).$$

Die Skalenkonstante gibt also die Wahrscheinlichkeit an, der mit Eins kodierten Gruppe anzugehören. Folglich könnte bei unbekannten Werten der Variablen  $x_q$ , ebenso wie bei  $\beta = 0$  eine Vorhersage (Diskrimination) der Gruppenzugehörigkeit *nur* nach den Gruppenwahrscheinlichkeiten erfolgen. Sind beide Gruppen zudem gleich wahrscheinlich, so ist eine begründete Entscheidung zwischen beiden Gruppen nicht mehr möglich, weil  $\beta_0 = 1/2$  wird.

Die Vorhersagbarkeit wächst hingegen mit der „Enge“ des Zusammenhangs zwischen  $y$  und den Variablen  $x_q$ , d.h., sofern wir von unveränderten Varianzen der beteiligten Variablen ausgehen, mit wachsender Höhe der Parameter  $\beta_q$ . Die Diskriminanzfunktion kann bei der hier vorliegenden zweiwertigen Gruppenvariablen  $y$  wegen

$$(2.42) \quad E(y|x) = P(y=1|x) \cdot 1 + P(y=0|x) \cdot 0 = P(y=1|x)$$

als bedingte Wahrscheinlichkeit, der Gruppe 1 anzugehören, interpretiert werden. Liegen die Parameter  $\beta_q \neq 0$  fest, so erhöhen die Absolutbeträge der Differenzen im Vektor  $x-E(x)$  die Sicherheit der Gruppenzuordnung.

Bei der Annahme einer linearen Diskriminanzfunktion ist zu beachten, daß der Wertebereich jeder quantitativen Variablen  $x_q$ ,  $q \in \{1,2,\dots,Q\}$ , deren zugehöriger Koeffizient  $\beta_q \neq 0$  ist, auf ein endliches Intervall beschränkt sein muß. Andernfalls ergäbe sich ein Widerspruch zu der Tatsache, daß die Werte von  $E(y|x)$  wegen der Kodierung mit Null und Eins als bedingte Wahrscheinlichkeiten zu interpretieren sind. Demnach läßt sich beispielsweise die Annahme der multivariaten Normalverteilung der  $x_q$  innerhalb der beiden Gruppen

nicht in Einklang bringen mit der Annahme, daß  $E(y | x)$  eine lineare Funktion im Sinne der Gleichung 2.1 ist, falls mindestens einer der Koeffizienten  $\beta_q, q \in \{1, 2, \dots, Q\}$ , ungleich Null ist.

Weiterführende Literatur: Eine einführende Darstellung in die Diskriminanzanalyse findet man z.B. bei Tatsuoka (1971) oder Marinell (1977). Bock (1975) zeigt darüber hinaus auch, wie man Nutzenüberlegungen mit der Diskriminanzanalyse verbinden kann, was für Zuordnungsprobleme insofern von Bedeutung ist, als verschiedene Arten von falschen Zuordnungen verschiedene große Schäden anrichten können.

### 2.3.4 Kontingenzanalyse

Von einer Kontingenzanalyse sprechen wir, wenn der statistische Zusammenhang zwischen qualitativen Variablen untersucht wird. Während wir im Abschnitt über die Varianzanalyse die Gruppenzugehörigkeit mit Variablen kodiert haben, deren Werte nur 1 und -1 sein konnten, so verwenden wir hier wie in der Diskriminanzanalyse Indikatorvariablen, die nur die Werte 1 und 0 annehmen. Dies führt zu besonders gut interpretierbaren Parametern.

Sind sowohl die unabhängigen Variablen  $x_q$  als auch die abhängige Variable  $y$  Kodiervariablen für jeweils zwei Kategorien, so reduziert sich die Untersuchung der Kovarianzen nach dem Modell der Gleichung 2.1 auf die Analyse der Wahrscheinlichkeit des Auftretens von Merkmalskombinationen. Exemplarisch betrachten wir hier den Fall einer  $2 \times 2 \times 2$  Kontingenzanalyse (siehe Abbildung 2.8) sowohl ohne als auch mit Interaktion.

Zur Interpretation der Parameter im kontingenzanalytischen Modell ohne Interaktion

$$(2.43) \quad E(y | x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

untersuchen wir wieder die Erwartungswerte von  $y$  bei gegebenem  $x = x$ . Nach der obigen Gleichung ergibt die Betrachtung des Falles  $x_1 = 0, x_2 = 0$

$$(2.44) \quad \beta_0 = E(y | x_1 = 0, x_2 = 0) = P(y = 1 | x_1 = 0, x_2 = 0),$$

die des Falles  $x_1 = 1, x_2 = 0$

$$(2.45) \quad \begin{aligned} \beta_1 &= E(y | x_1 = 1, x_2 = 0) - \beta_0 = \\ &= P(y = 1 | x_1 = 1, x_2 = 0) - P(y = 1 | x_1 = 0, x_2 = 0) \end{aligned}$$

und die des Falles  $x_1 = 0, x_2 = 1$

$$(2.46) \quad \begin{aligned} \beta_2 &= E(y | x_1 = 0, x_2 = 1) - \beta_0 = \\ &= P(y = 1 | x_1 = 0, x_2 = 1) - P(y = 1 | x_1 = 0, x_2 = 0). \end{aligned}$$

Gruppe	y	x <sub>1</sub>	x <sub>2</sub>	x <sub>1</sub>	.	x <sub>2</sub>
1	0	0	0	0		
2	1	0	0	0		
3	0	1	0	0		
4	1	1	0	0		
5	0	0	1	0		
6	1	0	1	0		
7	0	1	1	1		
8	1	1	1	1		

Abb. 2.8: Kodierung einer dreidimensionalen  $2 \times 2 \times 2$  Kontingenztafel für ein Modell mit Interaktion. Bei einer Kontingenzanalyse sind alle beteiligten Variablen qualitativ. Die unabhängigen Variablen  $x_1$  und  $x_2$  enthalten Information über die Gruppenzugehörigkeit einer Beobachtungseinheit. Bei einem Modell ohne Interaktion entfällt die Spalte für  $x_1 x_2$ . Eine Merkmalskombination liegt mit den Werten der Variablen y,  $x_1$  und  $x_2$  fest.

Der Parameter  $\beta_0$  ist also in einer  $2 \times 2 \times 2$  Kontingenzanalyse mit der Kodierung Null und Eins gleich der bedingten Wahrscheinlichkeit für „y = 1“ bei gegebenem  $x_1 = x_2 = 0$  (siehe Gleichung 2.44). Die Parameter  $\beta_1$  ist nach Gleichung 2.45 die Differenz der bedingten Wahrscheinlichkeiten von  $y = 1$  bei gegebenem  $x_1 = 1$  bzw.  $x_1 = 0$  auf der Stufe  $x_2 = 0$ , und der Parameter  $\beta_2$  ist nach Gleichung 2.46 die Differenz der bedingten Wahrscheinlichkeiten von  $y = 1$  bei gegebenem  $x_2 = 1$  bzw.  $x_2 = 0$  auf der Stufe  $x_1 = 0$ . Diese Interpretationen gelten sowohl für den Fall ohne als auch mit Interaktion.

Legen wir einen Zusammenhang zwischen y und den  $x_q$  zugrunde, bei dem eine Interaktion besteht, so können wir die Ausprägungskombinationen der Variablen  $x_1$  und  $x_2$  durch die Variable  $x_1 x_2$  erfassen (siehe die Spalte für  $x_1 x_2$  der Abbildung 2.8)

$$(2.47) \quad E(y | x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 \cdot x_2.$$

Als Identifikationsgleichungen für die Parameter  $\beta_1$  bis  $\beta_3$  dient auch hier die Gleichung 2.9, für  $\beta_0$  die Gleichung 2.11.

Die Interpretation der Parameter  $\beta_0$ ,  $\beta_1$  und  $\beta_2$  bleibt unverändert (siehe Gleichungen 2.44 bis 2.46); die für  $\beta_3$  ergibt sich aus der Gleichung 2.47 wie folgt:

$$(2.48) \quad \begin{aligned} \beta_3 &= E(y \mid x_1 = 1, x_2 = 1) - \beta_0 - \beta_1 - \beta_2 = \\ &= P(y = 1 \mid x_1 = 1, x_2 = 1) - P(y = 1 \mid x_1 = 1, x_2 = 0) - \\ &\quad - [P(y = 1 \mid x_1 = 0, x_2 = 1) - P(y = 1 \mid x_1 = 0, x_2 = 0)]. \end{aligned}$$

Eine Interaktion  $\beta_3 \neq 0$  liegt vor, wenn die Differenz der bedingten Wahrscheinlichkeiten  $P(y = 1 \mid x_2 = 1)$  und  $P(y = 1 \mid x_2 = 0)$  auf den beiden Stufen von  $x_1$  nicht gleich groß ist.

Als Beispiel für eine Kontingenzanalyse mit extremer Interaktion behandelt Moosbrugger (1980) das sogenannte Meehl'sche Paradoxon (Meehl, 1950).

Liegen für die Wahrscheinlichkeiten, einer Untergruppe anzugehören, Schätzungen vor, so können wir unter Beachtung entsprechender Voraussetzung (vgl. z.B. Bortz, 1977, S. 202) mit Signifikanztests ( $\chi^2$ -Tests) überprüfen, wie wahrscheinlich es ist, daß alle oder einzelne Parameter  $\beta_q$  in der Population gleich Null sind oder einen anderen hypothetischen festen Wert haben.

**Weiterführende Literatur:** Zur Kontingenzanalyse oder Analyse qualitativer Daten gibt es eine Vielzahl von unterschiedlichen Ansätzen, welche z.T. (wie etwa die „loglinearen Modelle“ oder die „Konfigurationsfrequenzanalyse“) keine Unterscheidung zwischen abhängigen und unabhängigen Variablen treffen. Die wichtigsten Ansätze scheinen uns durch folgende Autoren vertreten zu sein: Bishop, Fienberg & Holland (1975), Cox (1970), Goodman (1978), Krauth & Lienert (1973).

### 3. Modelle mit latenten Variablen

#### 3.1 Einleitung

Im Alltag erklärt man „gleichartiges“ Verhalten einer Person in verschiedenen Situationen oft, indem man ihr eine Eigenschaft zuschreibt, die für dieses Verhalten verantwortlich ist. Verhält sich jemand beispielsweise effektiv bei der Lösung von Problemen verschiedener Art, so schreibt man dies seiner hohen Intelligenz zu. Diese alltägliche Vorgehensweise kann man durch mathematische Modelle präzisieren. Das Prinzip dabei ist, daß man die Abhängigkeiten manifestierter oder beobachtbarer Variablen (das entspricht dem Verhalten in verschiedenen Situationen) durch eine oder mehrere latente, d.h. nicht

direkt beobachtbare Variablen (das entspricht der Eigenschaft) erklärt. Mit „Erklären“ meint man dabei, daß die Abhängigkeiten zwischen beobachtbaren Variablen verschwinden, wenn eine dahinterliegende latente Variable gefunden wird. Die latenten Variablen lassen sich dann als hypothetische Konstrukte (vgl. Herrmann, 1973 oder Mac Corquodale & Meehl, 1948) interpretieren.

Die latenten Eigenschaften können somit als Ursache für die Variation und Kovariation des beobachteten Verhaltens betrachtet werden (vgl. z.B. Bentler 1980). Beobachtbare Variablen, deren Kovariation auf eine gemeinsame latente Variable zurückzuführen ist, heißen auch *kongenerisch* (vgl. z.B. Jöreskog, 1971). Abbildung 3.1 zeigt ein einfaches Modell kongenerischer Variablen:

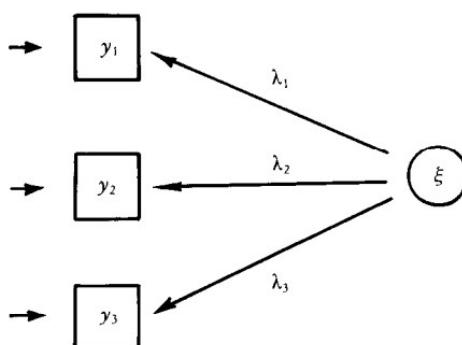


Abb. 3.1: Modell mit einer latenten Variablen  $\xi$  und drei beobachtbaren Variablen  $y_p$ , bei denen es sich in psychometrischen Anwendungen um drei Tests oder auch drei Testitems handeln kann. Die manifesten Variablen  $y_1$ ,  $y_2$  und  $y_3$  hängen mit den Gewichten  $\lambda_1$ ,  $\lambda_2$  und  $\lambda_3$  von der gemeinsamen latenten Variablen  $\xi$  sowie jeweils von den spezifischen Residualvariablen ab, welche durch unbeschriftete Pfeile gekennzeichnet sind.

In Kapitel 3.2 werden wir eine allgemeine *Theorie latenter Variablen* darstellen, auf deren Basis es möglich sein wird, eine Anzahl von Modellen, welche auf den ersten Blick durchaus verschieden erscheinen, als Spezialfalle dieser Theorie vorzustellen, weil sie sich nur durch die Art der beobachtbaren und latenten Variablen unterscheiden. Bei der linearen Faktorenanalyse<sup>7)</sup> mit einem Faktor sind sowohl die manifesten Variablen als auch die latente Variable (= Faktor) kontinuierlich. Beim „linear traceline model“ (siehe z.B. Lazarsfeld & Henry, 1968, S. 206 oder Fischer, 1974, S. 160) sind die manifesten Variablen zweiwertig und die latente Variable kontinuierlich. Beim „latent

<sup>7)</sup> Damit ist die Faktorenanalyse im Sinne von Spearman (1904, 1926), Thurstone (1947) und Jöreskog (1966) gemeint und nicht die Hauptkomponentenanalyse nach Pearson (1901) und Hotelling (1933). Die Unterschiedlichkeit beider Ansätze beschreiben z.B. Lawley & Maxwell, 1971, S. 2-4).

*profile model*" (siehe z.B. Gibson, 1959) sind die beobachtbaren Variablen kontinuierlich und die latente Variable repräsentiert nur die Zugehörigkeit zu einer latenten Klasse. Beim „*latent class model*“ (siehe z.B. Lazarsfeld & Henry, 1968, S. 46) sind die manifesten Variablen zweiwertig und die latente Variable repräsentiert ebenfalls nur die Zugehörigkeit zu den Klassen. Schließlich kann auch das „klassische latent-additive Testmodell“ (siehe Moosbrugger & Müller, 1980, 1981, 1982) in diesem Kontext abgehandelt werden.

Während allen vorgenannten Modellen eine lineare Variablencharakteristik zugrundeliegt, werden wir auch auf andere Gruppen von Modellen, nämlich solche mit polynomialem Variablencharakteristik (Lazarsfeld, 1950, McDonald 1967) sowie solche mit logistischer Variablencharakteristik (Rasch 1960, Birnbaum 1968) eingehen, die sich in formaler Hinsicht nur geringfügig von den linearen Modellen unterscheiden.

Die genannten Modelle können auch für mehrere latente Variablen verallgemeinert werden. Zur Einführung der wesentlichen Konzepte reicht jedoch der Fall einer einzigen latenten Variablen aus, welche die Abhängigkeiten zwischen den beobachtbaren Variablen erklärt. Auch in diesem dritten Abschnitt ist es nicht unser Ziel, einen umfassenden Überblick über Anwendungsmöglichkeiten, Probleme der Parameterschätzung und Hypothesenbeurteilung zu geben, sondern vielmehr die gemeinsamen theoretischen Grundlagen und Beziehungen herauszuarbeiten.

## 3.2 Eine allgemeine Theorie latenter Variablen

In diesem Abschnitt stellen wir zunächst die Annahmen der Theorie latenter Variablen und die wichtigsten Folgerungen zusammen. Danach untersuchen wir, wann und wie sich die unbekannten Parameter aus den statistischen Kennwerten manifestierter Variablen, nämlich aus Erwartungswerten, Varianzen und Kovarianzen, eindeutig bestimmen lassen.

### 3.2.1 Die Grundannahmen

Die formale Theorie latenter Variablen basiert auf zwei Grundannahmen, nämlich auf der bedingten stochastischen Unabhängigkeit der manifesten Variablen und auf der Art der Variablencharakteristik.

#### 3.2.1.1 Bedingte Unabhängigkeit

Der Grundgedanke von Modellen zur Erklärung der Abhängigkeiten beobachtbarer oder manifestierter Variablen  $y_p$ ,  $p \in \{1, 2, \dots, P\}$ , durch eine latente

Variable besteht darin, daß für jeden festen Wert der latenten Variablen  $\xi$  keine Abhängigkeit mehr zwischen den manifesten Variablen  $y_p$  besteht. Ein Beispiel soll dies erläutern: Handelt es sich bei den  $y_p$  um psychologische Testvariablen, welche miteinander korrelieren, so wird per Modellannahme erwartet, daß die Korrelationen verschwinden, wenn man die Testvariablen  $Vpn$  vorlegt, welche gleiche Ausprägungen der latenten Variablen  $\xi$  aufweisen. Für diese  $Vpn$  sollen die Testvariablen dann unabhängig sein.

Formal läßt sich dieser Gedanke wie folgt formulieren: Für alle nichtleeren Teilmengen  $J$  der Indexmenge  $\{1,2,\dots,p,\dots,P\}$  gelte (vgl. Anderson 1959, S. 11, Lord & Novick 1968, S. 543)

$$(3.1) \quad E \left( \prod_{p \in J} y_p \mid \xi \right) = \prod_{p \in J} E(y_p \mid \xi),$$

wobei  $\prod$  das Produktzeichen ist. Nach dieser Gleichung ist die bedingte Erwartung des Produkts der Variablen  $y_p$  gleich dem Produkt der bedingten Erwartungen der Variablen  $y_p$ . Zufallsvariablen  $y_p$ , welche diese Gleichung erfüllen, heißen bedingt linear stochastisch unabhängig.

Diese Bedingung sei für zwei Variablen  $y_p$ ,  $p \in \{1,2\}$  mit den Werten  $y_p = 1$  und  $y_p = 0$  veranschaulicht. Die bedingte Erwartung ist dann als bedingte Wahrscheinlichkeit interpretierbar, da  $E(y_p \mid \xi) = 1 \cdot P(y_p = 1 \mid \xi) + 0 \cdot P(y_p = 0 \mid \xi) = P(y_p = 1 \mid \xi)$  gilt. In Gleichung 3.1 erkennen wir dann das bekannte Multiplikationstheorem der Wahrscheinlichkeitsrechnung (vgl. z.B. Bortz, 1977, S. 71), welches besagt, daß zwei Ereignisse dann bedingt stochastisch unabhängig sind, wenn ihre Verbundwahrscheinlichkeit  $P(y_1 = 1 \mid \xi), (y_2 = 1 \mid \xi)] = P(y_1 = 1 \mid \xi) \cdot P(y_2 = 1 \mid \xi)$  als Produkt der bedingten Einzelwahrscheinlichkeiten darstellbar ist.

In der allgemeinen Theorie latenter Variablen machen wir ausdrücklich keine Verteilungsannahmen. Für die beteiligten Zufallsvariablen setzen wir jedoch voraus, daß sie eine endliche Erwartung haben.

### 3.2.1.2 Variablencharakteristische Funktion (VC-Funktion)

Die bedingte Erwartung  $E(y_p \mid \xi)$  nennen wir im vorliegenden Zusammenhang auch variablencharakteristische Funktion, abgekürzt VC-Funktion<sup>8</sup>). Sie gibt für jeden Wert  $\xi$  der latenten Variablen  $\xi$  an, welchen (bedingten) Erwartungswert  $E(y_p \mid \xi = \xi)$  die manifeste Variable  $y_p$ ,  $p \in \{1,2,\dots,P\}$ , hat. Die Verschiedenheit der hier behandelten Modelle zur Erklärung der Abhängigkeit

<sup>8)</sup> Andere Bezeichnungen sind: Variablen-, Itemcharakteristik, itemcharakteristische Funktion).

manifesten Variablen durch eine latente Variable besteht darin, von welchem Typ die variablencharakteristische Funktion  $E(y_p | \xi)$  ist.

a) Bei *linearer VC-Funktion* legt man den Typ von  $E(y_p | \xi)$  durch die Gleichung

$$(3.2) \quad E(y_p | \xi) = \lambda_{p0} + \lambda_p \xi$$

fest. Nach Gleichung 3.2 ist die bedingte Erwartung der manifesten Variablen  $y_p$  eine lineare Funktion der latenten Variablen  $\xi$ . In Abbildung 3.2 ist eine lineare VC-Funktion für den speziellen Fall kontinuierlicher latenter und manifesten Variablen  $\xi$  (siehe z.B. Abb. 3.7) auch die Variablencharakteristik  $E(y_p | \xi)$  diskret ist.

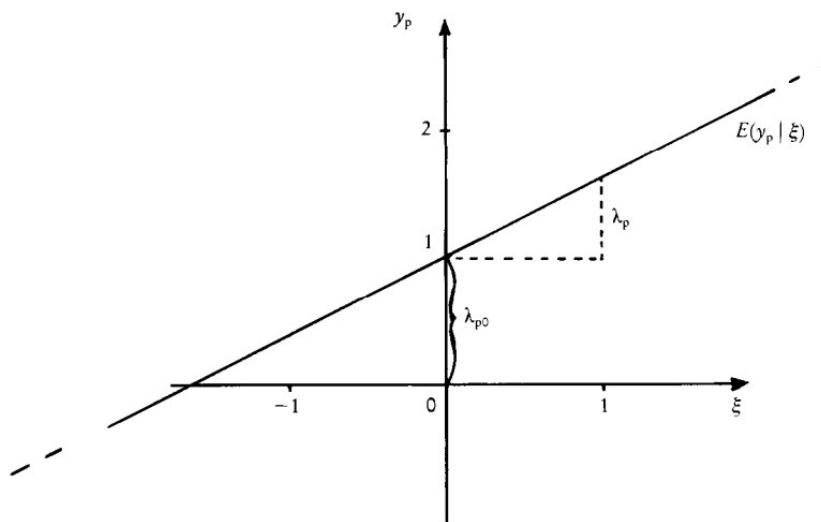


Abb. 3.2: Lineare VC-Funktion bei kontinuierlichen manifesten Variablen und einer kontinuierlichen latenten Variablen. Die Gerade  $E(y_p | \xi)$  gibt für jeden Wert  $\xi$  der latenten Variablen  $\xi$  den für  $y_p$  erwarteten Wert an. In diesem Fall wurden willkürlich  $\lambda_{p0} = 1.0$  und  $\lambda_p = 0.6$  angenommen.

b) Bei *polynomialer VC-Funktion* legt man den Typ von  $E(y_p | \xi)$  durch die Gleichung

$$(3.3) \quad E(y_p | \xi) = \lambda_{p0} + \lambda_{p1}\xi + \lambda_{p2}\xi^2 + \dots$$

fest. Die bedingte Erwartung der manifesten Variablen  $y_p$  ist dann eine kurvilineare Funktion der latenten Variablen  $\xi$ , wie sie in Abb. 3.3 für Polynome 2. Ordnung dargestellt ist:

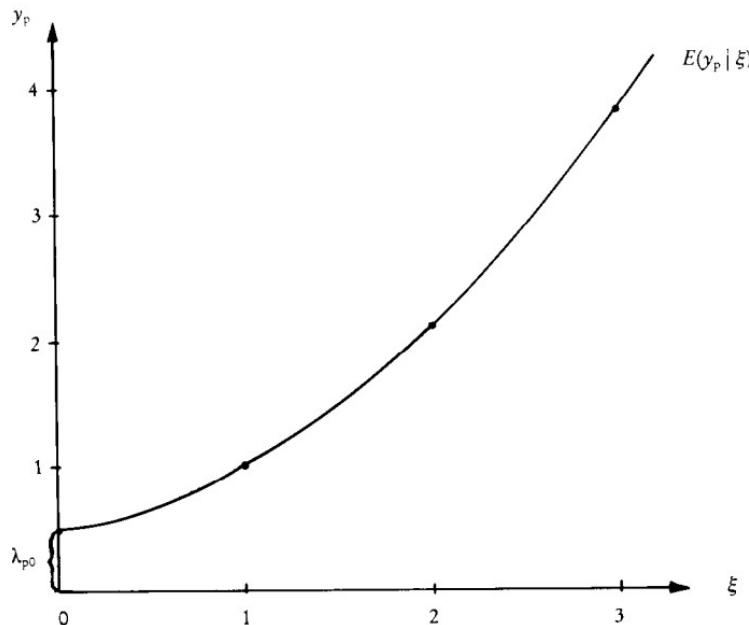


Abb. 3.3: Nichtlineare VC-Funktion bei kontinuierlichen manifesten Variablen und einer kontinuierlichen latenten Variablen in Form eines Polynoms 2. Ordnung, hier willkürlich mit  $\lambda_{p0} = 0.5$ ,  $\lambda_{p1} = 0.2$  und  $\lambda_{p2} = 0.3$ . Die Parabel  $E(y_p | \xi)$  gibt für den festen Wert  $\xi$  der latenten Variablen  $\xi$  den für  $y_p$  erwarteten Wert an.

c) Bei logistischer VC-Funktion legt man den Typ von  $E(y_p | \xi)$  durch die Gleichung

$$(3.4) \quad E(y_p | \xi) = \frac{\exp(\lambda_{p0} + \lambda_p \xi)}{1 + \exp(\lambda_{p0} + \lambda_p \xi)}$$

fest, wobei  $\exp(\lambda_{p0} + \lambda_p \xi) = e^{\lambda_{p0}} + \lambda_p \xi$  gilt mit  $e \cong 2.718$ , der Euler'schen Zahl. Die bedingte Erwartung der manifesten Variablen  $y_p$  ist dann eine logistische Funktion der latenten Variablen  $\xi$ . Der Zusammenhang ist in Abbildung 3.4, S. 31, dargestellt:

### 3.2.3 Die Residualvariablen

Bei Modellen mit einer latenten Variablen  $\xi$  ist der Zusammenhang zwischen  $\xi$  und den manifesten Variablen  $y_p$  nicht deterministisch, d.h. zwischen den  $y_p$ ,  $p \in \{1, 2, \dots, P\}$ , und deren bedingten Erwartungen  $E(y_p | \xi)$  wird in der Regel eine Differenz

$$(3.5) \quad \varepsilon_p := y_p - E(y_p | \xi),$$

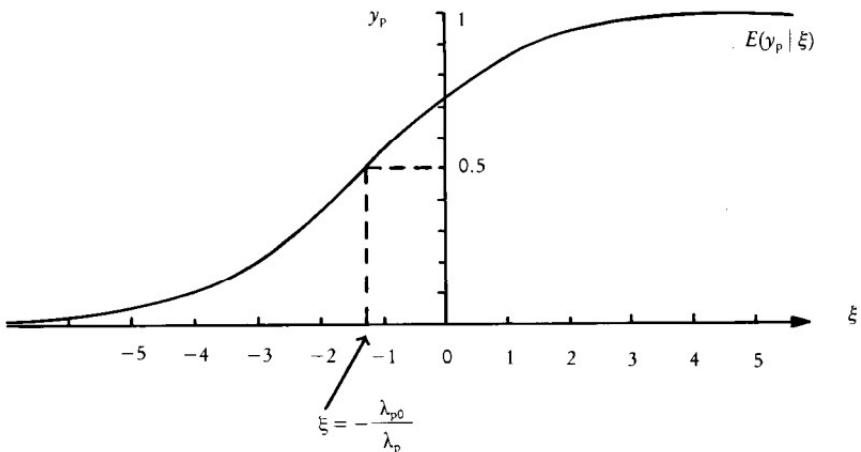


Abb. 3.4: Logistische VC-Funktion, hier willkürlich mit den Parametern  $\lambda_{p0} = 1.0$  und  $\lambda_p = 0.8$ . Der Parameter  $\lambda_p$  beschreibt die „Steilheit“ der VC-Funktion, der Parameter  $\lambda_{p0}$  ihre Lokation. Bei  $\lambda_{p0} + \lambda_p \xi = 0$ , d.h. bei  $\xi = -\lambda_{p0}/\lambda_p$  gilt immer  $E(y_p | \xi) = 1/2$ . Die bedingte Erwartung kann hier nur Werte zwischen Null und Eins annehmen. Auch bei diskreten manifesten Variablen  $y_p$  ist  $E(y_p | \xi)$  kontinuierlich.

bestehen, die wir als Residualvariablen bezeichnen. Diese haben die Eigenschaft, daß ihr bedingter Erwartungswert für alle Ausprägungen von  $\xi$  gleich Null ist:

$$(3.6) \quad E(\varepsilon_p | \xi) = 0$$

(siehe Anhang C, Regel 10). Für die Residualvariablen  $\varepsilon_p$  gelten die weiteren Eigenschaften

$$(3.7) \quad E(\varepsilon_p) = E[E(\varepsilon_p | \xi)] = 0,$$

(siehe Anhang C, Regel 1), und

$$(3.8) \quad C(\xi, \varepsilon_p) = C[\xi, y_p - E(y_p | \xi)] = 0$$

(siehe Anhang C, Regel 11). Nach Gleichung 3.7 ist also auch der unbedingte Erwartungswert der  $\varepsilon_p$  gleich Null, ebenso wie nach Gleichung 3.8 die Kovarianzen zwischen  $\xi$  und den  $\varepsilon_p$  Null sind.

Eine weitere Konsequenz aus der Modellannahme 3.1 ist, daß Residualvariablen  $\varepsilon_p$  paarweise nicht kovariieren

$$(3.9) \quad C(\varepsilon_p, \varepsilon_{p^*}) = 0, \text{ für alle } p, p^* \in \{1, 2, \dots, P\} \text{ mit } p \neq p^*.$$

### 3.2.4 Kovarianzmodellgleichungen und Identifikation

Im folgenden untersuchen wir für lineare VC-Funktionen, welche Folgerungen sich aus den Modellannahmen (Gleichungen 3.1 und 3.2) für die Modellgleichungen der Erwartungswerte, Varianzen und Kovarianzen der beobachtbaren Variablen  $y_p$  ergeben, und anschließend, ob und wie sich die Parameter  $\lambda_{p0}$ ,  $\lambda_p$ ,  $V(\xi)$  und  $V(\varepsilon)$  aus den Erwartungswerten, Varianzen und Kovarianzen der  $y_p$  eindeutig bestimmen lassen. Dies ist die Frage der Identifizierbarkeit. In manchen Fällen können, wie wir in Abschnitt 3.3.3 erwähnen, auch Modellgleichungen für Momente höherer Ordnung zur eindeutigen Bestimmung von Modellparametern beitragen.

Nach den Gleichungen 3.2 und 3.5 gilt für  $y_p$

$$(3.10) \quad y_p = \lambda_{p0} + \lambda_p \xi + \varepsilon_p.$$

Daraus und aus den Gleichungen 3.8 und 3.9 folgen die Modellgleichungen für die Kovarianzen der Variablen  $y_p$

$$\begin{aligned} (3.11) \quad C(y_p, y_{p^*}) &= C(\lambda_{p0} + \lambda_p \xi + \varepsilon_p, \lambda_{p^*0} + \lambda_{p^*} \xi + \varepsilon_{p^*}) = \\ &= \lambda_p C(\xi, \xi) \lambda_{p^*} + C(\varepsilon_p, \varepsilon_{p^*}) = \\ &= \begin{cases} \lambda_p \lambda_{p^*} V(\xi), & \text{für } p \neq p^* \\ \lambda_p^2 V(\xi) + V(\varepsilon_p), & \text{für } p = p^* \end{cases} \\ &\text{für alle } p, p^* \in \{1, 2, \dots, p\}. \end{aligned}$$

Für  $p = p^*$  geben die Gleichungen 3.11 die Modellgleichungen für die Varianz der beobachtbaren Variablen  $y_p$  an. Die durch die latente Variable  $\xi$  erklärte Varianz einer manifesten Variablen  $y_p$  ist gleich  $\lambda_p^2 V(\xi)$  und somit gleich  $\lambda_p^2$ , wenn  $\xi$  auf die Varianz  $V(\xi) = 1$  normiert wird.

Liegen drei beobachtbare Variablen  $y_p$  vor, wie in Abbildung 3.1 veranschaulicht, so implizieren die Annahmen 3.1 und 3.2 genau sechs Kovarianzmodellgleichungen (vgl. auch die „accounting equations“ von Lazarsfeld 1959, S. 504), welche die drei bekannten Varianzen und die drei bekannten Kovarianzen der beobachtbaren Variablen  $y_p$  mit den drei unbekannten Ladungen  $\lambda_p$  und den drei Fehlervarianzen  $V(\varepsilon_p)$  verbinden. Die Gleichungen 3.11 enthalten in diesem Fall genau  $P(P + 1)/2 = 6$  Modellgleichungen für die Kovarianzen, die gerade genügen, um eindeutig die sechs Parameter  $\lambda_1, \lambda_2, \lambda_3, V(\varepsilon_1), V(\varepsilon_2), V(\varepsilon_3)$  zu bestimmen, wenn wir entweder die Varianz der latenten Variablen  $\xi$  festlegen oder einem Parameter  $\lambda_p$  einen bestimmten festen Wert geben:

Wir leiten zunächst aus den Modellgleichungen 3.11 für die Kovarianzen der beobachtbaren Variablen die Identifikationsgleichung für  $\lambda_1$  ab. Dazu bilden wir durch dreimaliges Benutzen von Gleichung 3.11 für  $p \neq p^*$  den Ausdruck

$$(3.12) \quad \frac{C(y_1, y_2) \cdot C(y_1, y_3)}{C(y_2, y_3)} = \frac{\lambda_1 V(\xi) \lambda_2 \cdot \lambda_1 V(\xi) \lambda_3}{\lambda_2 V(\xi) \lambda_3} = \lambda_1^2 V(\xi).$$

Für den Koeffizienten  $\lambda_1$  erhalten wir demnach

$$(3.13) \quad \lambda_1 = \pm \sqrt{\frac{C(y_1, y_2) \cdot C(y_1, y_3)}{C(y_2, y_3) \cdot V(\xi)}}.$$

Wenn wir die Variablen  $y_p$  so definieren, daß sie positiv mit  $\xi$  kovariieren, so hat  $\lambda_1$  immer ein positives Vorzeichen. Da die Kovarianzen der manifesten Variablen  $y_p$  gegebene Größen sind, sieht man aus den Gleichungen 3.12 und 3.13, daß mit der Festlegung der Varianz  $V(\xi)$  der Koeffizient  $\lambda_1$  gegeben ist bzw. umgekehrt mit der Festlegung von  $\lambda_1$  auch  $V(\xi)$ . Da die Numerierung der Variablen  $y_p$  beliebig ist, kann man Gleichung 3.13 entsprechend für jede der Variablen  $y_p$ ,  $p \in \{1, 2, 3\}$  verwenden.

Wenn wir bedenken, daß wegen  $C(y_p, \xi) = C(\lambda_{p0} + \lambda_p \xi + \varepsilon_p, \xi) = \lambda_p V(\xi)$  für die Koeffizienten  $\lambda_p$

$$(3.14) \quad \lambda_p = \frac{C(y_p, \xi)}{V(\xi)}$$

gilt, so sehen wir, daß  $\lambda_p$  um so größer wird, je kleiner die Varianz von  $\xi$  festgelegt wird. Darüber hinaus ist zu beachten, daß  $\lambda_p > \lambda_{p^*}$  nur dann auf einen stärkeren Zusammenhang zwischen  $y_p$  und  $\xi$  gegenüber  $y_{p^*}$  und  $\xi$  hinweist, wenn  $y_p$  und  $y_{p^*}$  gleiche Varianzen haben.

Im Abschnitt 3.3.3 werden wir einen Fall kennenlernen, bei dem  $V(\xi)$  nicht beliebig festzulegen ist. In diesem Fall hat  $V(\xi)$  eine wichtige empirische Bedeutung, und wir werden sehen, daß durch Berücksichtigung der Modellstruktur für die zentralen Momente dritter Ordnung  $C(y_1, y_2, y_3)$  in diesem Fall auch die Varianz  $V(\xi)$  eindeutig bestimmt werden kann.

Aus den Gleichungen 3.11 (für  $p = p^*$ ) erhalten wir auch die Identifikationsgleichungen für die Fehlervarianzen  $V(\varepsilon_p)$ , wenn wir für  $\lambda_p^2 V(\xi)$  den linken Ausdruck aus Gleichung 3.12 einsetzen. Die folgende Gleichung gilt wieder für alle Variablen  $y_p$ , wenn man die Indices entsprechend umnumeriert:

$$(3.15) \quad V(\varepsilon_1) = V(y_1) - \frac{C(y_1, y_2) \cdot C(y_1, y_3)}{C(y_2, y_3)}$$

Sind die Parameter  $\lambda_p$  und  $V(\varepsilon_p)$  erst einmal eindeutig bestimmt, so kann man auch die Skalenkonstanten  $\lambda_{p0}$  unter Verwendung von Gleichung 3.10 aus

$$(3.16) \quad E(y_p) = \lambda_{p0} + \lambda_p E(\xi)$$

eindeutig bestimmen, wenn man  $E(\xi)$  festlegt. Gewöhnlich wird die Normierung  $E(\xi) = 0$  gewählt.

Allgemein erhält man bei einem Modell mit P kongenerischen Variablen genau  $P(P+1)/2$  Kovarianzgleichungen (siehe Gleichungen 3.11) mit  $2 \cdot P$  unbekannten Parametern, nämlich den  $\lambda_p$  und  $V(\varepsilon_p)$ , und außerdem P Gleichungen für die Erwartungswerte der  $y_p$  mit weiteren P unbekannten Parametern, nämlich den Skalenkonstanten  $\lambda_{p0}$ . Liefert ein Modell nur so viele Gleichungen für die Kovarianzen wie zu bestimmende Parameter, so kann das Modell zwar identifizierbar sein, es ist aber empirisch noch nicht überprüfbar, da es zu einer perfekten Anpassung des Modells an die Daten einer Stichprobe führen würde. Erst wenn  $P > 3$  ist, hat man mehr Gleichungen als Unbekannte, und man spricht von Überidentifikation. In diesem Fall können zumindest einige Parameter durch mehrere Gleichungen bestimmt werden und das Modell wird empirisch überprüfbar.

### 3.2.5 Beispiel

Ein fiktives Beispiel soll die oben dargestellte Theorie erläutern. Wir gehen davon aus, daß die folgenden Varianzen, Kovarianzen und Korrelationen zwischen den kongenerischen Variablen  $y_1$ ,  $y_2$  und  $y_3$  für eine bestimmte Population bekannt und in Tabelle 3.1 gesammelt sind. Dabei enthält die untere Dreiecksmatrix die Korrelationen, die obere die Kovarianzen und die Hauptdiagonale die Varianzen:

Tabelle 3.1: Korrelations- (untere Dreiecksmatrix) und Kovarianzmatrix (obere Dreiecksmatrix) fiktiver Daten.

	$y_1$	$y_2$	$y_3$
$y_1$	4.00	0.90	4.20
$y_2$	0.90	0.25	1.20
$y_3$	0.70	0.80	9.00

Die Parameter  $\lambda_p$  und  $V(\varepsilon_p)$ , die unter der Normierung  $V(\xi) = 1$  nach den Gleichungen 3.13 und 3.15 berechnet wurden, sind in Tabelle 3.2 gesammelt.

Wählen wir z.B. eine Personengruppe A mit der Ausprägung  $\xi_A$  in der latenten Variablen und eine Personengruppe B mit der um eine Einheit höheren Ausprägung  $\xi_B = \xi_A + 1$ , so erwarten wir nach Gleichung 3.2 für Personen der

Tabelle 3.2: Modellparameter für das fiktive Zahlenbeispiel der Tabelle 3.1.

	$\lambda_p$	$V(\varepsilon_p)$
$y_1$	1.77	0.87
$y_2$	0.50	0.00
$y_3$	2.37	3.38

Gruppe B verglichen mit Personen der Gruppe A in der manifesten Variablen  $y_1$  eine um 1.77, in  $y_2$  eine um 0.50 und in  $y_3$  eine um 2.37 Einheiten „höhere“ Merkmalsausprägung. Eine vergleichende Reihung der Gewichte  $\lambda_p$  hinsichtlich des Grades der Abhängigkeit der manifesten Variablen  $y_p$  von der latenten Variablen  $\xi$  ist erst möglich, wenn alle beobachteten Variablen auf die gleiche Varianz normiert werden.

Bis auf Rundungsfehler lässt sich mit diesen Daten und den Gleichungen 3.11 die Ausgangskovarianzmatrix exakt reproduzieren. Bei drei Variablen  $y_p$  lässt sich diese perfekte Anpassung des Modells an die Daten mit jeder beliebigen Korrelationsmatrix erzielen, in der alle Eintragungen von Null verschieden sind. Erst bei vier beobachteten Variablen  $y_p$  wäre dies nicht zwangsläufig der Fall, womit das Modell dann empirisch überprüfbar wäre. Dabei ist die inhaltlich wichtigste Frage, ob eine latente Variable zur Erklärung der beobachteten Zusammenhänge ausreicht.

### 3.2.6 Modell paralleler Variablen

Führt man noch restriktivere Annahmen ein, als sie in der formalen Theorie latenter Variablen mit linearer VC-Funktion (Gleichungen 3.1 und 3.2) gemacht wurden, erhält man das Modell paralleler Variablen (vgl. hierzu den Begriff „parallele Tests“ in der klassischen Testtheorie, z.B. in Fischer, 1974, S. 34). Beim Modell paralleler Variablen werden zusätzlich die Gleichungen

$$(3.17) \quad \lambda_{p0} = \lambda_0; \lambda_p^2 = \lambda^2; V(\varepsilon_p) = V(\varepsilon), \text{ für alle } p \in \{1, 2, \dots, P\},$$

angenommen. Parallele Variablen haben demnach gleiche Erwartungswerte (das folgt aus den Gleichungen 3.16 und 3.17), gleiche erklärte Varianzen  $\lambda^2 V(\xi)$  (siehe Gleichungen 3.11 für  $p = p^*$ ) und gleiche Fehlervarianzen. Parallelle Variablen sind in diesem Sinn austauschbar, denn sie messen dieselbe latente Eigenschaft auf der gleichen Skala gleich gut. Kongenerische Variablen, die nicht zugleich parallel sind, messen hingegen dieselbe Eigenschaft auf verschiedenen Skalen verschieden gut.

### 3.2.7 Zusammenfassende Bemerkungen

Die in diesem Abschnitt dargestellte allgemeine formale Theorie liegt den in den folgenden Abschnitten zu behandelnden Modellen zugrunde. Sie besteht aus zwei Grundannahmen, der bedingten stochastischen Unabhängigkeit (siehe Gleichung 3.1) und der Art der Variablencharakteristik (siehe Gleichungen 3.2 bis 3.4). Diese beiden Annahmen genügen zur eindeutigen Bestimmung der Modellparameter (siehe für lineare VC-Funktion die Gleichungen 3.13 und 3.15) sofern Varianzen und Kovarianzen von genügend vielen beobachtbaren Variablen vorhegen und die Varianz der latenten Variablen festgelegt wird. Bei restriktiveren Annahmen über die Modellparameter genügen für deren eindeutige Bestimmung bereits zwei beobachtbare kongenerische Variablen, die wir dann auch „parallel“ nennen.

*Weiterführende Literatur:* Die Theorie latenter Variablen ist zunächst nur in Ansätzen behandelt worden (siehe z.B. Anderson 1959, Lord & Novick 1968). Eine monographische Abhandlung liefert Harnerle (1982).

## 3.3 Anwendungen der allgemeinen Theorie latenter Variablen

### 3.3.1 Faktorenanalyse

Als erste Anwendung der im Abschnitt 3.2 dargestellten Theorie behandeln wir die lineare Faktorenanalyse mit einem Faktor (einer latenten Variablen).

Von der Faktorenanalyse in dem hier verwendeten engeren Sinn (Spearman 1904, 1926, Thurstone 1947, Jöreskog 1966) muß die „Hauptkomponentenanalyse“ (Pearson 1901, Hotelling 1933) unterschieden werden. Während es Ziel der Faktorenanalyse ist, die Kovarianzmatrix von kongenerischen Variablen durch eine geringe Anzahl von Faktoren (latenten Variablen) zu erklären, beschränkt sich die Hauptkomponentenmethode darauf, eine Kovarianzmatrix von Variablen durch orthogonale Komponenten, welche auch Faktoren genannt werden, aber dem theoretischen Anspruch latenter Variablen meist nicht genügen, zu beschreiben (vgl. z.B. Lawley & Maxwell, 1971, S. 2-4). Die Hoffnung in der bisherigen Forschung, exploratorisch mit der häufig angewandten Hauptkomponentenmethode geeignete Theorien „automatisch“ zu finden, konnte sich deshalb nur kaum erfüllen. Doch seit Jöreskog (1967, 1969, 1973) rechentechnische Probleme in der Faktorenanalyse gelöst hat, ist auch die Überprüfung von Theorien mit konfirmatorischen Faktorenanalysen praktisch leicht durchführbar (siehe Sörbom & Jöreskog, 1976, Jöreskog & Sörbom, 1978).

In der Faktorenanalyse werden sowohl die latente Variable  $\xi$  (der Faktor) als auch die beobachtbaren Variablen  $y_p$  als kontinuierlich angenommen. Dann ist

auch die lineare Variablencharakteristik  $E(y_p | \xi)$  kontinuierlich (siehe Abbildung 3.5). Sie kann als Regressionsgerade zwischen der jeweiligen Variablen  $y_p$  und der latenten Variablen  $\xi$  angesehen werden. Die Werte  $\xi$  von  $\xi$  heißen in der Faktorenanalyse „factor scores“.

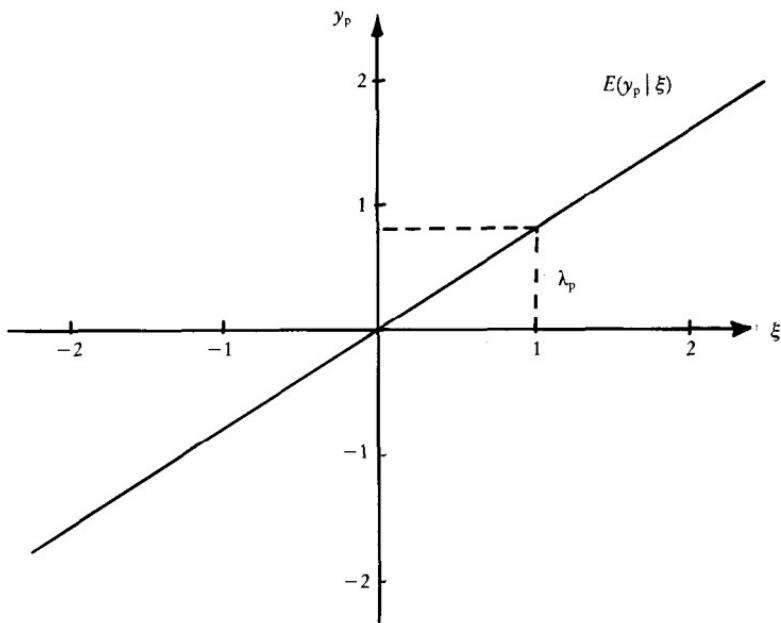


Abb. 3.5: Faktorenanalytisches Modell mit einem Faktor als Spezialfall der allgemeinen Theorie latenter Variablen mit linearer VC-Funktion entsprechend Gleichung 3.2 sowie kontinuierlichen manifesten und kontinuierlichen latenten Variablen. Die Gerade  $E(y_p | \xi)$  gibt für jeden Wert  $\xi$  (factor score) der latenten Variablen  $\xi$  den für  $y_p$  erwarteten Wert an. In der Faktorenanalyse sind die Normierungen  $E(y_p) = E(\xi) = 0$  üblich, wodurch die Skalenkonstanten  $\lambda_{p0} = 0$  sind. Die Ladung  $\lambda_p$  wurde hier willkürlich mit  $\lambda_p = 0.8$  angenommen.

Die Gewichte  $\lambda_p$  werden in der Faktorenanalyse „Ladungen“ der Variablen  $y_p$  auf  $\xi$  genannt. Ist die Varianz  $V(\xi)$  auf Eins normiert, so ist  $\lambda_p^2$  die „Kommunalität“ der Variablen  $y_p$ . Die Gleichungen 3.11 zeigen, daß es sich dabei um die durch  $\xi$  erklärte Varianz von  $y_p$  handelt, also um  $V[E(y_p | \xi)]$ .

Als klassischen Anwendungsfall können wir Spearman's Generalfaktor-Theorie betrachten (Spearman, 1904).

*Weiterführende Literatur:* Als deutschsprachige Lehrbücher seien Pawlik (1968), Revenstorf (1976, 1980) und Überla (1971) genannt. Revensdorf (1980), Anderson & Rubin (1956) sowie Lawley & Maxwell (1971) gehen ausführlicher auf die statistischen Probleme der Faktorenanalyse ein, als die Monographien von Mulaik (1972), Harman (1967) oder Horst (1965), welche ebenso wie die oben angeführten deutschsprachigen Autoren mehr Platz der Hauptkomponentenanalyse widmen. Die nichtlineare Faktorenanalyse (entsprechend Gl. 3.3) wird z.B. von McDonald (1967) behandelt.

### 3.3.2 „Linear traceline model“

In diesem Abschnitt untersuchen wir, welche Besonderheiten sich für die Interpretation einiger Gleichungen und Parameter ergeben, wenn man zusätzlich zu den Modellannahmen der Gleichungen 3.1 und 3.2 festlegt, daß die latente Variable kontinuierlich ist, die manifesten Variablen aber zweiwertig sind.

Lazarsfeld (vgl. Lazarsfeld & Henry, 1968, S. 206) hat ein Modell eingeführt, das von einer linearen Beziehung zwischen der kontinuierlichen latenten Variablen  $\xi$  und den zweiwertigen beobachtbaren Variablen  $y_p$  ausgeht, das „linear traceline model“. Wir zeigen daher in diesem Abschnitt zunächst, daß das Lazarsfeld'sche Modell mit linearer Variablencharakteristik („linear traceline“) ein Spezialfall des allgemeinen Modells latenter Variablen ist, wenn man von zweiwertigen manifesten Variablen und einer kontinuierlichen latenten Variablen ausgeht.

In den Gleichungen 3.1 und 3.2 sind die Annahmen des allgemeinen linearen Modells latenter Variablen formuliert. Die zusätzlichen Annahmen des „linear traceline model“ bestehen in der notwendigen (vgl. dazu Wottawa, 1979, S. 49) Einschränkung von  $\xi$  auf das Intervall  $[\xi_1, \xi_2]$  und in der Festlegung, daß  $\xi$  kontinuierlich ist, die  $y_p$  aber nur die Werte 0 bzw. 1 annehmen können. Wir untersuchen nun, welche Folgerungen sich aus diesen Festlegungen ergeben:

Die erste Besonderheit betrifft die lineare VC-Funktion, für die in diesem Fall gilt:

$$(3.18) \quad E(y_p | \xi) = P(y_p = 1 | \xi) = \lambda_{p0} + \lambda_p \xi.$$

Demnach ist die bedingte Wahrscheinlichkeit, daß  $y_p$  den Wert 1 unter  $\xi = \xi$  annimmt, eine lineare Funktion der latenten Variablen  $\xi$  (siehe Abbildung 3.6). Hier wird deutlich, daß das allgemeine Modell latenter Variablen zu einem „probabilistischen Modell“ wird, wenn  $y_p$  nur zwei Werte annehmen kann, welche mit 0 und 1 kodiert werden (vgl. auch Abschnitt 3.3.4). Der einzige Unterschied zwischen dem faktorenanalytischen und dem probabilistischen Modell in Gleichung 3.18 besteht darin, daß die Werte der bedingten

Erwartung  $E(y_p | \xi)$  im letzteren Fall als bedingte Wahrscheinlichkeiten interpretiert werden können.

Bei psychometrischen Tests, in denen  $\xi$  eine latente Fähigkeit repräsentiert, kann bei der Kodierung  $y_p = 0$  für „Nichtlösung“ und  $y_p = 1$  für „Lösung“ einer Aufgabe p die Konstante  $\lambda_{p0}$  als Leichtigkeit der Aufgabe p interpretiert werden. Die unbedingten Erwartungswerte der beobachtbaren Variablen

$$(3.19) \quad E(y_p) = E(\lambda_{p0} + \lambda_p \xi + \varepsilon_p) = \lambda_{p0} + \lambda_p E(\xi),$$

vereinfachen sich nämlich bei der Normierung  $E(\xi) = 0$  zu

$$(3.20) \quad E(y_p) = \lambda_{p0} = P(y_p = 1).$$

Man sieht, daß  $\lambda_{p0}$  hier die Lösungswahrscheinlichkeit der Aufgabe p darstellt, welche in der klassischen Testtheorie (vgl. Lienert 1969, S. 87, od. Moosbrugger, 1981 b, S. 27) als Aufgabenschwierigkeit bezeichnet wird.

Der Parameter  $\lambda_p$  zeigt die „Trennschärfe“ einer Aufgabe an. Je größer  $\lambda_p$  ist, desto größer der Zusammenhang (siehe Gleichung 3.14) zwischen Fähigkeit und Lösung dieser Aufgabe, und desto weniger hängt das Lösen der Aufgabe von anderen, nicht erfaßten Umständen ab.

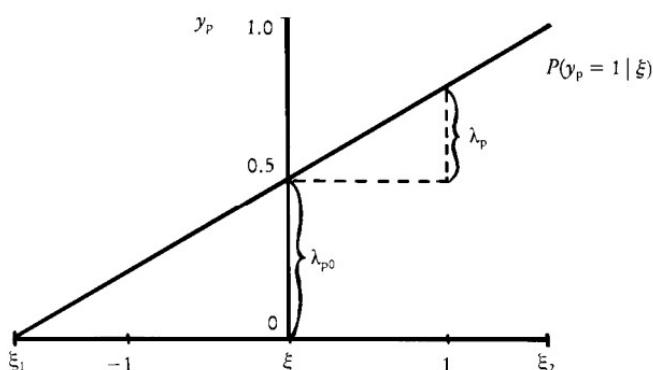


Abb. 3.6: Linear traceline model. Die bedingte Wahrscheinlichkeit, daß die Variable  $y_p$  den Wert 1 annimmt, ist eine lineare Funktion der Ausprägung der latenten Variablen  $\xi$ . Man sieht, daß die bedingte Erwartung (Wahrscheinlichkeit) auch bei diskreten  $y_p$  kontinuierlich ist. Wurde  $\xi$  normiert mit  $E(\xi) = 0$ , dann ist  $\lambda_{p0}$  die unbedingte Wahrscheinlichkeit für „ $y_p = 1$ “. Bei Fähigkeitstests kann  $\lambda_{p0}$  deshalb als „Leichtigkeit“ interpretiert werden. Der Parameter  $\lambda_p$  gibt den Wert an, um den die Wahrscheinlichkeit für „ $y_p = 1$ “ ansteigt, wenn sich  $\xi$  um eine Einheit erhöht. Zugleich ist  $\lambda_p$  bei  $V(\xi) = 1$  die Kovarianz zwischen  $y_p$  und  $\xi$  (siehe Gleichung 3.14). In diesem Fall wurden willkürlich  $\lambda_{p0} = 0.5$  und  $\lambda_p = 0.3$  angenommen.

**Weiterführende Literatur:** Lazarsfeld & Henry (1968) schreiben an mehreren Stellen ausführlicher über das „linear traceline model“. Auch Fischer (1974, S. 160ff.) geht auf dieses Modell ein. Ein Anwendungsbeispiel findet man in Lazarsfeld (1959). Auch die polynomiale latent structure analysis von Lazarsfeld (1950) kann als Spezialfall der allgemeinen Theorie gemäß Gleichung 3.1 und Gleichung 3.3 dargestellt werden.

### 3.3.3 „Latent profile-“ und „latent class model“

Wir betrachten in diesem Abschnitt den Fall, daß es sich bei der latenten Variablen  $\xi$  um eine zweiwertige Kodiervariable handelt, die nur die Werte „Null“ oder „Eins“ annehmen kann und somit die Zugehörigkeit zu einer von zwei Klassen anzeigt. Die manifesten Variablen  $y_p$ ,  $p \in \{1, 2, \dots, P\}$ , hingegen können kontinuierlich oder ebenfalls zweiwertig sein. Gibson (1959) nennt den ersten Fall „latent profile model“, auf den auch Lazarsfeld & Henry (1968, S. 228) sowie McDonald (1967, S. 58) eingehen. Der Fall, daß sowohl die latente Variable als auch die manifesten Variablen dichotom sind, ist unter der Bezeichnung „latent class model“ bekannt (vgl. Lazarsfeld & Henry, 1968, S. 17). Beide Modelle sind Spezialfälle der allgemeinen Theorie, die wir in Abschnitt 3.2 dargestellt haben, d.h. die beiden Modelle sind charakterisiert durch die Gleichungen 3.1 und 3.2 sowie die Festlegungen, daß beim „latent profile model“  $y_p$  kontinuierlich für alle  $p \in \{1, 2, \dots, P\}$ , und  $\xi$  zweiwertig mit den Werten 0 oder 1 sind, wohingegen beim „latent class model“ sowohl  $y_p$  als auch  $\xi$  zweiwertig ist mit den Werten 0 oder 1, für alle  $p \in \{1, 2, \dots, P\}$ .

Wie man durch Berechnung der bedingten Erwartungswerte sehen kann, vereinfacht sich in diesen Fällen der Parameter  $\lambda_{p0}$  zu

$$(3.21) \quad E(y_p | \xi = 0) = E(\lambda_{p0} + \lambda_p \xi + \varepsilon_p | \xi = 0) = \lambda_{p0},$$

(siehe Gleichung 3.6 und Anhang A, Regel 3), und im Fall manifeste zweiwertiger Variablen gilt

$$(3.22) \quad P(y_p = 1 | \xi = 0) = \lambda_{p0}.$$

Demnach ist also  $\lambda_{p0}$  der bedingte Erwartungswert von  $y_p$  bzw. die bedingte Wahrscheinlichkeit für „ $y_p = 1$ “ im Fall  $\xi = 0$ . Auf entsprechende Weise finden wir für  $\lambda_p$  im Fall kontinuierlicher manifeste Variablen

$$(3.23) \quad E(y_p | \xi = 1) = E(\lambda_{p0} + \lambda_p \xi + \varepsilon_p | \xi = 1) = \lambda_{p0} + \lambda_p,$$

(siehe Gleichung 3.5 und Anhang A, Regel 3), und durch Einsetzen der Gleichung 3.21 erhalten wir die Gleichung

$$(3.24) \quad \lambda_p = E(y_p | \xi = 1) - E(y_p | \xi = 0),$$

die sich im Fall zweiwertiger manifester Variablen vereinfacht zu

$$(3.25) \quad \lambda_p = P(y_p = 1 | \xi = 1) - P(y_p = 1 | \xi = 0).$$

Nach diesen Gleichungen ist  $\lambda_p$  im latent profile model gleich der Differenz der bedingten Erwartungswerte von  $y_p$  unter „ $\xi = 1$ “ und „ $\xi = 0$ “ (siehe Abb. 3.7) bzw. im latent class model der Differenz der bedingten Wahrscheinlichkeiten für „ $y_p = 1$ “ unter „ $\xi = 1$ “ und „ $\xi = 0$ “ (siehe Abb. 3.8). Ist  $\lambda_p$  gleich Null, so differenzieren die beiden latenten Klassen nicht hinsichtlich den betreffenden manifesten Variablen  $y_p$ .

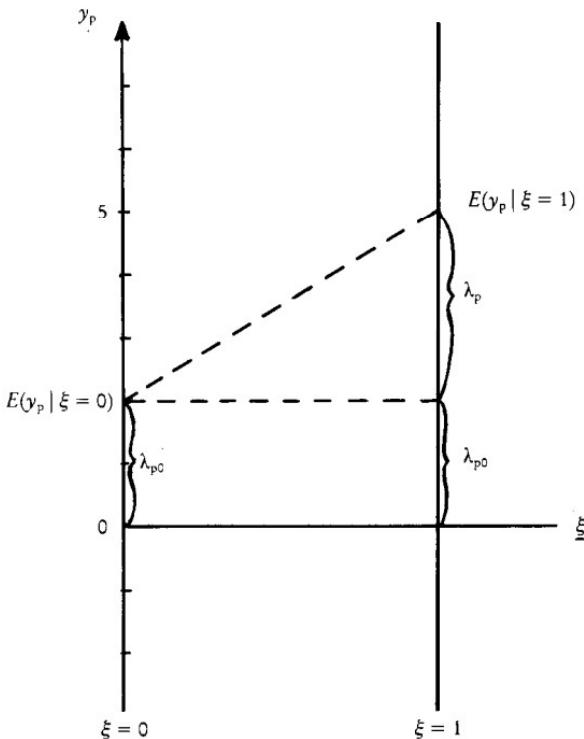


Abb. 3.7: Latent profile model als Spezialfall der allgemeinen Theorie latenter Variablen mit linearer VC-Funktion gemäß Gleichung 3.2 sowie kontinuierlichen manifesten und dichtomer 0/1 kodierter latenten Variablen. Die bedingte Erwartung hat dann nur zwei Werte, nämlich  $\lambda_{p0}$  (hier 2) auf der Stufe  $\xi = 0$  und  $\lambda_{p0} + \lambda_p$  (hier  $2 + 3 = 5$ ) auf der Stufe  $\xi = 1$ .

Die grundlegende Modellvorstellung der bedingten stochastischen Unabhängigkeit der  $y_p$  (siehe Gleichung 3.1) bedeutet hier, daß innerhalb einer latenten Klasse die manifesten Variablen nicht korrelieren und auch in keiner Abhängigkeit höherer Ordnung stehen (vgl. Beispiel 3.2.5).

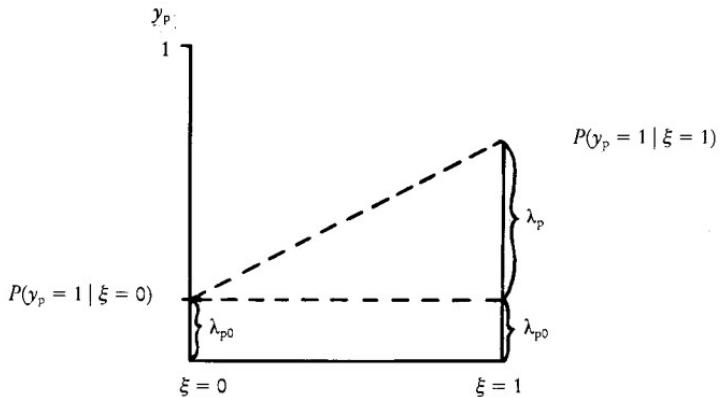


Abb. 3.8: Latent class model als Spezialfall der allgemeinen Theorie latenter Variablen mit linearer VC-Funktion gemäß Gleichung 3.2 sowie dichotomen 0/1 kodierten manifesten und dichotomer 0/1 kodierter latenter Variablen. Die bedingte Erwartung hat dann nur zwei Werte, nämlich die Wahrscheinlichkeiten für  $y_p = 1$  auf der Stufe  $\xi = 0$ , welche nach Gleichung 3.22 gleich  $\lambda_{p0}$  (hier 0.2) ist, sowie die Wahrscheinlichkeit für  $y_p = 1$  auf der Stufe  $\xi = 1$  (hier 0.7). Die Differenz beider Wahrscheinlichkeiten beschreibt nach Gleichung 3.25 der Parameter  $\lambda_p$  (hier 0.5).

Bei einem Modell, bei dem eine dichotome latente Variable mit Null bzw. Eins kodiert wird, kann man die Standardisierung  $E(\xi) = 0$  und  $V(\xi) = 1$  nicht einführen, da diese beiden Kennwerte hier eine besondere Bedeutung haben, denn

$$(3.26) \quad E(\xi) = P(\xi=1)$$

und

$$(3.27) \quad V(\xi) = P(\xi=1) \cdot [1 - P(\xi=1)]$$

(siehe Anhang A, Gleichung A.1).

Die Wahrscheinlichkeit, einer der beiden Klassen anzugehören, kann theoriegeleitet à priori festgelegt oder empirisch ermittelt werden, wenn man die Modellstrukturgleichung für das zentrale Moment dritter Ordnung zur Identifikation der Klassenwahrscheinlichkeiten  $P(\xi=1)$  und  $P(\xi=0)$  benutzt. Näheres siehe Steyer & Moosbrugger, 1979, S. 38-41 sowie Lazarsfeld & Henry, 1968, S. 42. Ein illustratives Beispiel findet man in Lazarsfeld & Henry (1968, S. 27ff.).

Weiterführende Literatur: Das „latent class model“ wird ausführlich in Lazarsfeld & Henry (1968) behandelt. Einen kurzen Überblick gibt Madansky (1978). Mit Problemen der Parameterschätzung befassen sich Goodman (1974, 1979) und Formann (1978, 1980).

### 3.3.4 Logistische Modelle

Bei dem in Abschnitt 3.3.2 dargestellten „linear traceline model“ mußten wir  $\xi$  auf ein bestimmtes Intervall einschränken. Andernfalls könnte die bedingte Erwartung auch Werte größer als Eins oder kleiner als Null annahmen (vgl. auch Wottawa, 1980, S. 49), welche mit der Wahrscheinlichkeitsinterpretation unvereinbar wären. Beim logistischen Testmodell von Birnbaum (1968) und dessen Spezialfall, dem dichotomen logistischen Testmodell von Rasch (1960) wird diesem Problem vorgebeugt, indem außer der bedingten stochastischen Unabhängigkeit (siehe Gleichung 3.1) anstelle der linearen VC-Funktion (siehe Gleichung 3.2) die logistische VC-Funktion (siehe Gleichung 3.4) angenommen wird. Bei dichotomen manifesten Variablen  $y_p$  mit den Werten 0 bzw. 1 werden die Testmodelle „probabilistisch“ und als bedingte Erwartung erhalten wir folgende bedingte Wahrscheinlichkeit (vgl. Gleichung 3.4):

$$(3.27) \quad E(y_p | \xi) = P(y_p = 1 | \xi) = \frac{\exp(\lambda_{p0} + \lambda_p \xi)}{1 + \exp(\lambda_{p0} + \lambda_p \xi)}$$

In der logistischen Modellgleichung 3.27 kommen zwei Parameter vor, die spezifisch für die Variable  $y_p$  sind;  $\lambda_p$  beschreibt die „Steilheit“ der VC-Funktion,  $\lambda_{p0}$  ihre Lokalisation entlang der  $\xi$ -Achse (vgl. Abbildung 3.4). In psychometrischen Anwendungen können wir  $\lambda_p$  als Maß für die Trennschärfe („discriminating power“ sensu Birnbaum 1968, S. 409) einer Variablen auffassen und  $\lambda_{p0}$  als „Leichtigkeit“. Anstelle von  $\lambda_{p0}$  verwendet Birnbaum (1968, S. 409) einen Schwierigkeitsparameter  $\lambda_{p0}^*$ , welcher nach Ausklammern von  $\lambda_p$  in Gleichung 3.27 und Setzen von  $\lambda_{p0}/\lambda_p = -\lambda_{p0}^*$  gefunden wird. Die Modellgleichung 3.27 lautet dann

$$(3.28) \quad E(y_p | \xi) = P(y_p = 1 | \xi) = \frac{\exp[\lambda_p(\xi - \lambda_{p0}^*)]}{1 + \exp[\lambda_p(\xi - \lambda_{p0}^*)]}$$

Der Schwierigkeitsparameter  $\lambda_{p0}^*$  ermöglicht einen direkten Vergleich mit dem Personenparameter  $\xi$ , welcher bei Leistungstests zu besonders einfachen Parameterinterpretationen führt: Bezeichnet nämlich  $\xi$  die „Fähigkeit“ einer Person in der latenten Variablen  $\xi$  und  $\lambda_{p0}^*$  die Schwierigkeit der Aufgabe, so beträgt wegen  $\exp(0) = 1$  die Lösungswahrscheinlichkeit  $P(y_p = 1 | \xi = \xi) = 1/2$ , wenn  $\xi = \lambda_{p0}^*$  ist, nämlich genau dann, wenn die Fähigkeit der Person ebenso groß wie die Schwierigkeit der Aufgabe ist. Bei positiv gewählten  $\lambda_p$  gilt  $P(y_p = 1 | \xi) > 1/2$ , wenn  $\xi > \lambda_{p0}^*$  und umgekehrt.

Ein interessanter Spezialfall des logistischen Testmodells und damit der allgemeinen Theorie latenter Variablen ist das dichotome logistische Testmodell von Rasch (1960), bei dem alle Parameter  $\lambda_p$ ,  $p \in \{1, 2, \dots, P\}$  gleich Eins gesetzt werden. Das bedeutet, daß alle Variablen  $y_p$  die latente Variable  $\xi$  „gleich gut“ messen sollen, aber auf im allgemeinen unterschiedlichem Schwierigkeitsniveau. Demnach verlaufen die VC-Funktionen der verschiedenen  $y_p$  parallel, aber jeweils um die Differenz ihrer  $\lambda_p$  entlang der  $\xi$ -Achse verschoben. Abbildung 3.9 illustriert diesen Sachverhalt:

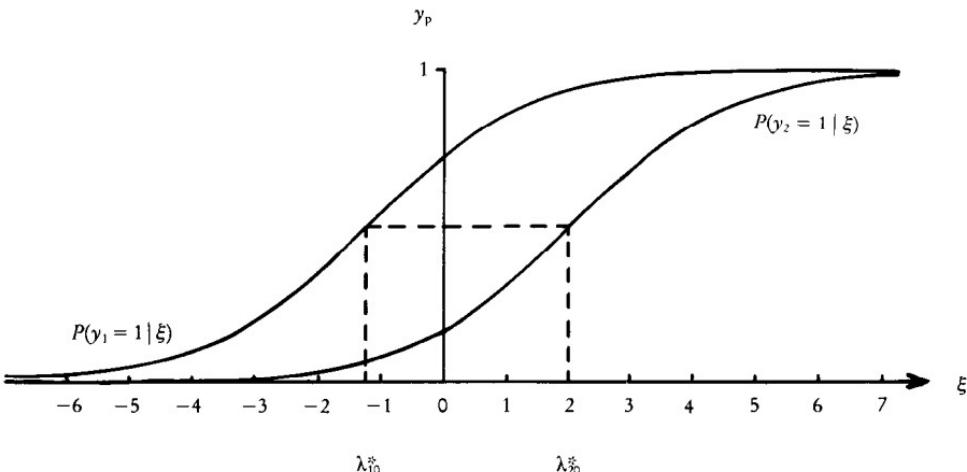


Abb. 3.9: Dichotomes logistisches Testmodell nach Rasch (1960) als Spezialfall der allgemeinen Theorie latenter Variablen mit logistischer VC-Funktion entsprechend Gleichung 3.4 sowie dichotomen 0/1 kodierten manifesten und kontinuierlicher latenten Variablen. Die Kurven  $P(y_p = 1 | \xi)$  geben für jeden Wert  $\xi$  der latenten Variablen  $\xi$  die Wahrscheinlichkeit für die Ausprägung  $y_p = 1$  an; bei psychometrischen Anwendungen ist das die „Lösungswahrscheinlichkeit“ der Aufgabe  $p$ . Die VC-Funktionen der Aufgaben  $p = 1$  und  $p = 2$  haben gleiche „Trennschärfe“ ( $\lambda_1 = \lambda_2 = 1$ ), aber unterschiedliche Schwierigkeiten  $\lambda_{10}^* = -1.25$  und  $\lambda_{20}^* = 2.0$ . Ihre Schwierigkeitsdifferenz beträgt  $\lambda_{20}^* - \lambda_{10}^* = 3.25$ . Für Personen mit  $\xi = \xi_1 = -1.25$  erhalten wir bei Aufgabe 1 die Lösungswahrscheinlichkeit 1/2 wie auch für Personen mit  $\xi = \xi_2 = 2.0$  bei Aufgabe 2. Mit der Wahl konstanter  $\lambda_p$  ist auch die Varianz der latenten Variablen festgelegt.

Die bei Modellgültigkeit gleichen  $\lambda_p$  für alle  $p \in \{1, 2, \dots, P\}$  Aufgaben ermöglichen „spezifisch objektive“, d.h. aufgaben- bzw. personenstichprobenunabhängige, Vergleiche (siehe z.B. Fischer, 1978, S. 299) zwischen Personen- bzw. Aufgabenparameter.

*Weiterführende Literatur:* Als Weiterführung in deutscher Sprache können Fischer (1968, 1974), Rost & Spada (1978) sowie Wottawa (1980) angeführt werden. In englischer Sprache sei außer Birnbaum (1968) und Lord & Novick (1968) auch das Heft 2 des Bandes 14 (1977) des Journal of Educational Measurement genannt, in dem man neben einer Arbeit von Hambleton & Cook (1977) mehrere ein- und weiterführende Beiträge zum Modell von Rasch und dem von Birnbaum findet. Ein Übersichtsreferat gibt Fischer (1978).

### 3.3.5 Klassisches latent-additives Testmodell

Ein weiteres psychometrisches Testmodell, welches sich als Erweiterung der klassischen Testtheorie versteht, nämlich das „Klassische latent-additive Testmodell“ (KLA-Modell, Moosbrugger & Müller 1981, 1982, Müller 1980) läßt sich als Spezialfall der allgemeinen Theorie latenter Variablen darstellen: die manifesten Variablen  $y_p$  sind intervallskalierte Reaktionsvariablen, welche durch geeignete streng monotone Transformationen aus beobachtbaren kontinuierlichen Testvariablen  $x_p$  gewonnen werden. Der Zusammenhang zwischen den  $y_p$  und der kontinuierlichen latenten Variablen  $\xi$  wird mit einer linearen VC-Funktion gemäß Gleichung 3.2 beschrieben, wobei für alle Aufgaben  $\lambda_{p0} = 1$  gesetzt wird:

$$(3.29) \quad E(y_p | \xi) = \lambda_{p0} + 1 \cdot \xi.$$

Ähnlich wie im Birnbaum-Modell (vgl. Abschnitt 3.3.4) wird anstelle des Leichtigkeitsparameters ein Schwierigkeitsparameter  $\lambda_{p0}^* = -\lambda_{p0}$  benutzt, welcher wegen

$$(3.30) \quad E(\xi | y_p = 0) = -\lambda_{p0} = \lambda_{p0}^*$$

angibt (siehe Abbildung 3.10), in welchem Abstand vom Nullpunkt die VC-Funktion die  $\xi$ -Achse schneidet. Die KLA-Modellgleichung lautet dann in den hier verwendeten Ausdrücken

$$(3.31) \quad E(y_p | \xi) = \xi - \lambda_{p0}^*.$$

Der Personenparameter  $\xi$  und der Aufgabenparameter  $\lambda_{p0}^*$  können wegen Gleichung 3.30 auf einer Skala verglichen werden.

Gilt  $\xi > \lambda_{p0}^*$  ist also die Merkmalsausprägung  $\xi$  einer Person größer als die „Schwierigkeit“ der Aufgabe p, so wird eine positive Reaktion erwartet und umgekehrt eine negative Reaktion.

Wegen der bei Modellgültigkeit gleichen  $\lambda_p$  für alle  $p \in \{1, 2, \dots, P\}$  Aufgaben und der dann parallelen VC-Funktionen (vgl. Abbildung 3.10) weist auch das KLA-Modell die von Fischer (1974, S. 407) für das Rasch-Modell (siehe Abschnitt 3.3.4) hervorgehobene Modelleigenschaft „spezifische Objektivität“

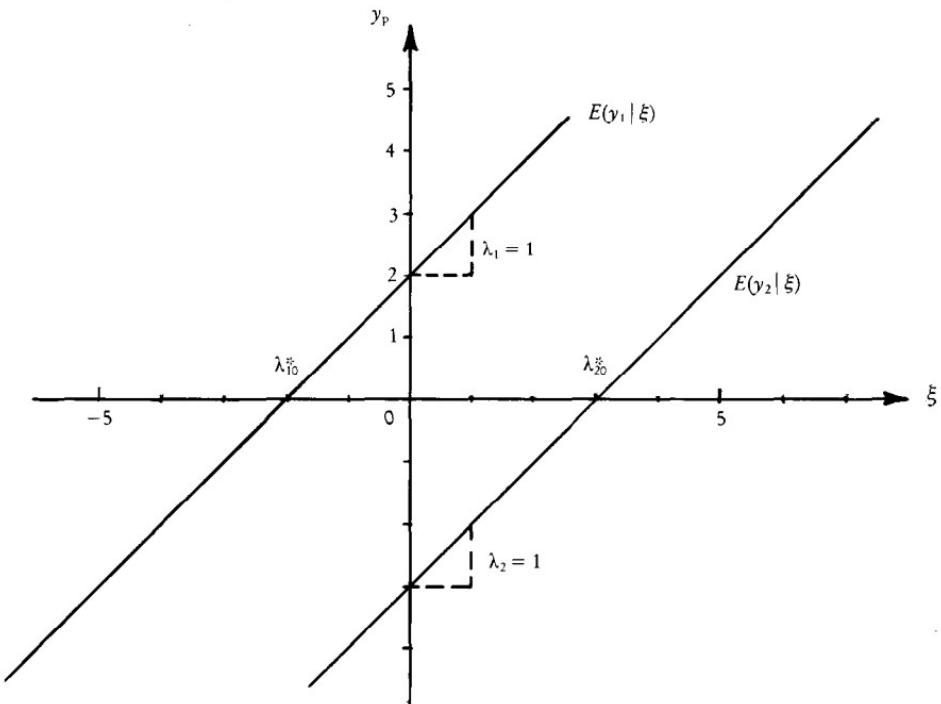


Abb. 3.10: Klassisches latent-additives Testmodell (Moosbrugger & Müller, 1981, 1982) als Spezialfall der allgemeinen Theorie latenter Variablen mit linearer VC-Funktion entsprechend Gleichung 3.2 sowie kontinuierlichen manifesten und kontinuierlichen latenten Variablen. Die Geraden  $E(y_p | \xi)$  gemäß Gleichung 3.31 geben bei psychometrischen Anwendungen für jeden Wert der latenten Variablen  $\xi$  an, welche Reaktion  $y_p$  erwartet wird. In jedem Item werden von jenen Personen positive Reaktionen erwartet, deren latente Merkmalsausprägung größer ist als die entsprechende Itemschwierigkeit, z.B. wenn im 1. Item  $\xi > \lambda_{10}^* = -2$  gilt bzw. im 2. Item  $\xi > \lambda_{20}^* = 3$ . Da die VC-Funktionen aller Aufgaben gleiche Steigungen  $\lambda_p = 1$  aufweisen, sind „spezifisch objektive Vergleiche“ zwischen Personen und Aufgaben möglich.

auf; es erlaubt, Schwierigkeitsunterschiede zwischen Aufgaben mit beliebigen Personen und Merkmalsunterschiede zwischen Personen mit beliebigen Items zu beurteilen.

**Weiterführende Literatur:** Das KLA-Modell wurde von Moosbrugger & Müller (1980) hinsichtlich seiner wesentlichen Eigenschaften mit der klassischen Testtheorie und dem dichotomen logistischen Testmodell (siehe auch Abschnitt 3.3.4) verglichen. Zur Parameterschätzung siehe Müller (1980); über geeignete Computeralgorithmen berichten Müller & Moosbrugger (1980). Mit empirischen Anwendungen befassen sich Moosbrugger (1982), Ruch & Hehl (1982) sowie Tarnai (1981).

### 3.4 Zusammenfassende Bemerkungen

In diesem dritten Abschnitt haben wir eine formale Theorie vorgestellt, die beschreiben kann, wie beobachtbares Verhalten mit psychologischen Begriffen, Eigenschaften oder Konstrukten in Beziehung steht. Erst bei Vorliegen einer Theorie über die konkreten Beziehungen bestimmter psychologischer Konstrukte mit bestimmten beobachtbaren Variablen wird es sinnvoll, von diesen Konstrukten, Eigenschaften oder Begriffen zu sprechen.

Eine mathematisch formulierte Theorie über den Zusammenhang zwischen beobachtbarem Verhalten und den psychologischen Konstrukten hat den Vorteil größerer Präzision und leichterer Überprüfbarkeit.

Die hier behandelte Theorie latenter Variablen basiert auf den Annahmen der bedingten stochastischen Unabhängigkeit und der Art der Variablencharakteristik. Dabei ist es unerheblich, ob die beteiligten Variablen kontinuierlich oder nur zweiwertig sind. Die Einschränkung auf eine einzige latente Variable wurde nur der Einfachheit halber gemacht. Theorien mit mehreren latenten Variablen erfordern jedoch einen größeren mathematischen Aufwand, als wir in diesem Überblick betreiben wollten.

## 4. Anhänge

### 4.1 Anhang A:

*Regeln für Erwartungswerte (vgl. Hays, 1973, S. 871ff.)*

Der *Erwartungswert* einer stochastischen Variablen oder Zufallsvariablen kann als eine mathematisch exakte Formulierung des umgangssprachlichen Begriffes des Durchschnittswertes angesehen werden. Im Fall einer diskreten Zufallsvariablen  $x$  mit den Werten  $x_n$  kann man den Erwartungswert  $E(x)$  folgendermaßen definieren:

$$(A.1) \quad E(x) := \sum_n x_n \cdot P(x = x_n).$$

Der Erwartungswert  $E(x)$  ist demnach die Summe der mit ihren Auftretenswahrscheinlichkeiten gewichteten Werte  $x_n$  der stochastischen Variablen  $x$ . Im Fall, daß die stochastische Variable  $N$  verschiedene Werte hat und jeder dieser Werte mit der Wahrscheinlichkeit  $1/N$  auftritt, erhält die Gleichung A.1 die Form

$$(A.2) \quad E(x) = \sum_{n=1}^N x_n \cdot \frac{1}{N} = \frac{1}{N} \sum_{n=1}^N x_n.$$

Im Fall, daß  $x$  kontinuierlich ist, definiert man  $E(x)$  mit Hilfe des Integrals (siehe Hays, 1973, S. 871 oder Bauer, 1974, S. 138, Steyer 1982, in diesem Band).

Der Erwartungswertvektor eines stochastischen Spaltenvektors  $x = (x_1, \dots, x_Q)'$  ist dann einfach definiert durch

$$(A.3) \quad E(x) := \begin{bmatrix} E(x_1) \\ \vdots \\ E(x_Q) \end{bmatrix}$$

Die wichtigsten Regeln für diese Konzepte sind im folgenden angegeben:

*Regel 1:* Wenn  $\alpha$  ein Vektor konstanter reeller Zahlen ist, dann gilt

$$E(\alpha) = \alpha.$$

*Regel 2:* Wenn  $\alpha'$  ein Zeilenvektor konstanter reeller Zahlen und  $x$  ein stochastischer Vektor mit dem Erwartungswertvektor  $E(x)$  ist, dann gilt  
 $E(\alpha'x) = \alpha'E(x).$

*Regel 3:* Wenn  $\alpha'$  und  $\beta'$  Zeilenvektoren konstanter reeller Zahlen und  $x$  und  $y$  stochastische Vektoren mit den Erwartungswertvektoren  $E(x)$  und  $E(y)$  sind, dann gilt  
 $E(\alpha'x + \beta'y) = \alpha'E(x) + \beta'E(y).$

Diese Regeln gelten entsprechend auch für skalare Größen, welche als Vektoren mit einer Komponente aufgefaßt werden können.

## 4.2 Anhang B:

### Regeln für Varianzen und Kovarianzen

Die *Kovarianz* zweier stochastischer Variablen oder Zufallsvariablen ist ein Parameter, der die Enge ihres linearen statistischen Zusammenhangs widerspiegelt, d.h. sie gibt an, wie stark die beiden Variablen gemeinsam variieren. Die Kovarianz zweier stochastischer Variablen  $x$  und  $y$ , die wir mit  $C(x,y)$  symbolisieren, ist folgendermaßen definiert:

$$(B.1) \quad C(x,y) := E[[x - E(x)] \cdot [y - E(y)]].$$

Die Verallgemeinerung für mehrere stochastische Variablen ist die *Kovarianzmatrix*. Sind  $x$  bzw.  $y$  stochastische Spaltenvektoren mit  $Q$  bzw.  $P$  Komponenten, so ist deren Kovarianzmatrix  $C(x,y)$  wie folgt definiert:

$$(B.2) \quad C(x,y) := E[(x - E(x))(y - E(y))'] = \\ = \begin{bmatrix} C(x_1, y_1) & C(x_1, y_2) & \dots & C(x_1, y_p) & \dots & C(x_1, y_p) \\ C(x_2, y_1) & C(x_2, y_2) & \dots & C(x_2, y_p) & \dots & C(x_2, y_p) \\ \vdots & \vdots & & \vdots & & \vdots \\ C(x_q, y_1) & C(x_q, y_2) & \dots & C(x_q, y_p) & \dots & C(x_q, y_p) \\ \vdots & \vdots & & \vdots & & \vdots \\ C(x_Q, y_1) & C(x_Q, y_2) & \dots & C(x_Q, y_p) & \dots & C(x_Q, y_p) \end{bmatrix}$$

Ein Kovarianzvektor ergibt sich, wenn einer der beiden stochastischen Vektoren nur eine Komponente enthält. Ist z.B. Q=1, so reduziert sich die Kovarianzmatrix C(x,y) auf den Kovarianzvektor

$$(B.3) \quad C(x,y) = [C(x,y_1), C(x,y_2), \dots, C(x,y_p), \dots, C(x,y_p)].$$

Ist hingegen P = 1, so reduziert sie sich auf den Kovarianzvektor

$$(B.4) \quad C(x,y) = \begin{bmatrix} C(x_1, y) \\ C(x_2, y) \\ \vdots \\ C(x_q, y) \\ \vdots \\ C(x_Q, y) \end{bmatrix}$$

Man beachte, daß  $C(x,y) = C(y,x)'$  und  $C(y,x) = C(x,y)'$  gelten.

Im folgenden sind einige Regeln für diese Konzepte aufgeführt, die auch für die Varianz einer stochastischen Variablen von Bedeutung sind, da diese durch

$$(B.5) \quad V(x) := C(x,x)$$

definiert ist. Bei allen stochastischen Variablen setzen wir im folgenden voraus, daß sie einen endlichen Erwartungswert haben.

**Regel 1:** Sind  $\alpha$  ein Vektor von reellen Konstanten und x ein stochastischer Vektor, dann gilt

$$C(\alpha, x) = \mathbf{0}.$$

**Regel 2:** Wenn  $A_1, A_2, B_1, B_2$  Matrizen konstanter reeller Zahlen sind und  $x_1, x_2, y_1, y_2$  stochastische Vektoren mit entsprechenden Anzahlen von Komponenten, dann gilt

$$\begin{aligned} C(A_1 x_1 + A_2 x_2, B_1 y_1 + B_2 y_2) &= \\ &= A_1 C(x_1, y_1) B_1' + A_1 C(x_1, y_2) B_2' + A_2 C(x_2, y_1) B_1' + A_2 C(x_2, y_2) B_2' \end{aligned}$$

Diese Regeln gelten entsprechend auch für skalare Größen, welche als Vektoren mit einer Komponente aufgefaßt werden können.

## 4.3 Anhang C

### *Regeln für bedingte Erwartungen*

In diesem Anhang sind einige Rechenregeln für bedingte Erwartungen (vgl. Steyer, Moosbrugger & Ullrich 1981) zusammengestellt. Dabei sind  $x$ ,  $y$  und  $z$  stochastische Variablen und  $x$ ,  $y$  und  $z$  stochastische Vektoren auf dem gleichen Wahrscheinlichkeitsraum sowie  $\alpha$ ,  $\beta$  reellwertige Konstanten und  $\alpha$ ,  $\beta$  Vektoren reellwertiger Konstanten.

Die *bedingte Erwartung* von  $y$  unter  $x$ ,

$$E(y|x) = E(y|x_1, \dots, x_Q)$$

ist eine stochastische Variable auf dem gemeinsamen Wahrscheinlichkeitsraum von  $x$  und  $y$ , deren Werte  $E(y|x=x)$  die Erwartungswerte von  $y$  unter der Bedingung sind, daß der stochastische Vektor  $x$  einen bestimmten festen Wertevektor  $x$  angenommen hat (vgl. Feller, 1971<sup>2</sup>, S. 162). Für *bedingte Erwartungswerte*  $E(y, x=x)$  sind sinngemäß die gleichen Regeln wie für unbedingte Erwartungswerte gültig.

Der Vektor der bedingten Erwartungen ist definiert durch

$$(C.1) \quad E(y|x) := [E(y_1|x), \dots, E(y_P|x)]'.$$

Für *bedingte Erwartungen* gelten folgende *Regeln*, die z.T. der Arbeit von Zimmermann (1975) entnommen sind. Die meisten davon gelten „nur“ mit Wahrscheinlichkeit 1. Auf diese Differenzierung verzichten wir jedoch in diesem mehr an den Praktiker gerichteten Artikel.

*Regel 1:* Der Erwartungswertvektor des Vektors der bedingten Erwartung von  $y$  unter  $x$  ist gleich dem Erwartungswertvektor von  $y$ :  
 $E[E(y|x)] = E(y).$

*Regel 2:* Die bedingte Erwartung von  $x_q$ ,  $q \in \{1, \dots, Q\}$ , unter  $x = (x_1, \dots, x_Q)'$  ist gleich  $x_q$ :  
 $E(x_q|x_1, \dots, x_Q) = x_q.$

*Regel 3:* Die bedingte Erwartung eines reellwertigen Konstantenvektors  $\alpha$  unter  $x$  ist gleich  $\alpha$ :  
 $E(\alpha|x) = \alpha.$

*Regel 4:* Sind  $\alpha$  und  $\beta$  reellwertige Konstantenvektoren sowie  $x$  und  $y$  stochastische Vektoren, so gilt für die bedingte Erwartung von  $\alpha'x + \beta'y$  unter  $z$   
 $E(\alpha'x + \beta'y|z) = \alpha'E(x|z) + \beta'E(y|z).$

*Regel 5:* Alle Komponenten von  $x^*$  seien auch in  $x$  enthalten. Dann gilt: Die bedingte Erwartung unter  $x^*$  der bedingten Erwartung von  $y$  unter  $x$  ist gleich der bedingten Erwartung von  $y$  unter  $x^*$

$$E[E(y|x) | x^*] = E(y|x^*)$$

*Regel 6:* Die bedingte Erwartung des stochastischen Vektors  $x_q \cdot y$  unter  $x_1, \dots, x_Q$  ist gleich dem Produkt von  $x_q$  mit der bedingten Erwartung von  $y$  unter  $x_1, \dots, x_Q$ :

$$E[x_q \cdot y | x_1, \dots, x_Q] = x_q \cdot E(y | x_1, \dots, x_Q).$$

*Regel 7:* Der Erwartungswert der bedingten Erwartung der stochastischen Variablen

$\prod_{p=1}^P y_p := y_1 \cdot \dots \cdot y_P$  unter  $x$  ist gleich dem

Erwartungswert von  $\prod_{p=1}^P y_p$ :

$$E \left[ E \left( \prod_{p=1}^P y_p | x \right) \right] = E \left( \prod_{p=1}^P y_p \right).$$

*Regel 8:* Der Erwartungswert des stochastischen Vektors  $x_q \cdot E(y|x)$  ist gleich dem Erwartungswertvektor von  $x_q \cdot y$ :

$$E[x_q \cdot E(y | x_1, \dots, x_Q)] = E(x_q \cdot y)$$

*Regel 9:* Der Erwartungswert des stochastischen Vektors  $y - E(y|x)$  ist gleich dem Nullvektor:

$$E[y - E(y|x)] = 0.$$

*Regel 10:* Der Vektor der bedingten Erwartungen des stochastischen Vektors  $[y - E(y|x)]$  unter  $x$  ist gleich dem Nullvektor:

$$E[[y - E(y|x)] | x] = 0.$$

*Regel 11:* Der Kovarianzvektor von  $x_q$  und  $y - E(y | x_1, \dots, x_Q)$  ist gleich dem Nullvektor

$$C[x_q, y - E(y | x_1, \dots, x_Q)] = 0'.$$

*Regel 12:* Der Erwartungswert des Produkts der bedingten Erwartungen zweier Zufallsvariablen  $z$  und  $y$ , beide unter  $x$ , ist gleich dem Erwartungswert des Produkts der stochastischen Variablen  $z$  mit der bedingten Erwartung von  $y$  unter  $x$  und umgekehrt.

$$E[E(z|x) \cdot E(y|x)] = E[z \cdot E(y|x)] = E[E(z|x) \cdot y].$$

*Regel 13:* Wenn  $\alpha$  eine reellwertige Konstante sowie  $x_1, \dots, x_Q$  und  $y$  stochastische Variablen sind, dann gilt:

$$E(y | x_1, \dots, x_Q) = \alpha \Rightarrow C(y, x_q) = 0, \\ \text{für jedes } q \in \{1, 2, \dots, Q\}.$$

Diese Regeln gelten entsprechend auch für skalare Größen, welche als Vektoren mit einer Komponente aufgefaßt werden können.

## Literatur

- Adrain. 1818. Investigation of the figure of the earth and of the gravity in different latitudes. *Transactions of the American Philosophical Society* 1. 119-135.
- Anderson, T. W. & Rubin, H. 1956. Statistical inference in factor analysis. *Proceedings of the 3<sup>rd</sup> Berkeley Symposium on Mathematical Statistics and Probability*, 5, 111-150.
- Anderson, T. W. 1959. Some scaling models and estimation procedures in the latent class model. In: Grenander, U. (Ed.): *Probability and statistics, The Harold Cramér Volume*. New York: Wiley, 9-38.
- Anderson, T. W. 1978. Classification and discrimination. In: Kruskal, W. H. & Tanur, J. M. (Eds): *International Encyclopedia of Statistics*. New York: Free Press.
- Bandemer, H. et al. 1977. *Theorie und Anwendung der optimalen Versuchsplanung I*. Berlin: Akademie-Verlag.
- Bartussek, D. 1970. Eine Methode zur Bestimmung von Moderatoreffekten. *Diagnostica*, 16, 57-76.
- Bauer, H. 1974. *Wahrscheinlichkeitstheorie und Grundzüge der Meßtheorie*. Berlin: De Gruyter.
- Bentler, P. M. 1980. Multivariate analysis with latent variables: causal modeling. *Annual Review of Psychology*, 31, 419-456.
- Birnbaum, A. 1968. Some latent trait models and their use in inferring an examinee's ability. In: Lord & Novick (Eds): *Statistical theories of mental test scores*. London: Addison-Wesley, 377-479.
- Bishop, Y. M. M. & Fienberg, S. E. & Holland, P. W. 1975. *Discrete multivariate analysis: Theory and practice*. Cambridge (Mass.): MIT-Press.
- Bock, R. D. 1975. *Multivariate statistical methods in behavioral research*. New York: McGraw-Hill.
- Bortz, J. 1977. *Lehrbuch der Statistik*. Berlin: Springer.
- Conger, A. J. & Jackson, D. N. 1972. Suppressor variables, prediction and the interpretation of psychological relationship. *Educational and Psychological Measurement*, 32, 579-599.

- Conger, A. J. 1974. A revised definition for suppressor variables; a guide to their identification and interpretation. *Educational and Psychological Measurement*, 34, 35-46.
- Cox, D. R. 1970. *The analysis of binary data*. London: Methuen.
- Cramer, E. M. & Nicewander, W. A. 1979. Some symmetric, invariant measures of multivariate association. *Psychometrika*, 44, 43-54.
- Darlington, R. B. 1968. Multiple regression in psychological research and practice. *Psychological Bulletin*, 69, 161-182.
- David, F. N. 1978. Galton, Francis: In: Kruskal, W. H. & Tanur, J. M. (Eds): *International Encyclopedia of Statistics*. New York: Free Press.
- Darper, N. R. & Smith, H. 1966. *Applied regression analysis*. New York: Wiley.
- Edwards, A. L. 1976. *An introduction to linear regression and correlation*. San Francisco: Freeman.
- Feller, W. 1971<sup>2</sup>. *An introduction to probability theory and its applications*. New York: Wiley.
- Finn, J. D. 1974. *A general model for multivariate analysis*. New York: Holt, Rinehart & Winston.
- Fischer, G. H. 1968. *Psychologische Testtheorie*. Bern: Huber.
- Fischer, G. H. 1974. *Einführung in die Theorie psychologischer Tests*. Bern: Huber.
- Fischer, G. H. 1978. Probabilistic test models and their applications. *German Journal of Psychology* 2, 298-319.
- Fisher, R. A. 1925. *Statistical methods for research workers*. London: Oliver & Boyd.
- Fisher, R. A. 1936: The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7, 179-188.
- Formann, A. K. 1978. A note on Parameter estimation for Lazarsfeld's latent class analysis. *Psychometrika* 43, 123-126.
- Formann, A. K. 1980. Neuere Verfahren der Parameterschätzung in der Latent-Class-Analyse. *Zeitschrift für Differentielle und Diagnostische Psychologie* 1, 107-116.
- Gaensslen, H. & Schubö, W. 1973. *Einfache und komplexe statistische Analyse*. München: Reinhardt.
- Galton, F. 1877. Typical laws of heredity. Royal Institution of Great Britain. *Proceedings* 8, 282-301.
- Galton, F. 1889. *Natural inheritance*. London, New York: Macmillan.
- Gibson, W. A. 1959. Three multivariate models: Factor analysis, latent structure analysis and latent profile analysis. *Psychometrika* 24, 229-252.
- Goodman, L. A. 1974. Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* 61, 215-231.
- Goodman, L. A. 1978. *Analyzing qualitative/categorial data*. London: Addison-Wesley.

- Goodman, L. A. 1979. On the estimation of Parameters in latent structure analysis. *Psychometrika* 44, 123-128.
- Guilford, J. P. 1967. The structure of human intellect. New York: Mc Graw Hill.
- Hambleton, R. K. & Cook, L. L. 1977. Latent trait models and their use in the analysis of educational test data. *Journal of Educational Measurement*, 14, 75-96.
- Harnerle, A. 1982. Latent-Trait-Modelle. Die meßtheoretische und multivariate statistische Analyse kategorialer Reaktionsmodelle. Weinheim und Basel: Beltz.
- Harman, H. H. 1967. Modern factor analysis. Chicago: The University of Chicago Press.
- Harter, H. L. 1974: The method of least squares and some alternatives - Part I, II. *International Statistical Review* 42, 147-174; 235-264.
- Harter, H. L. 1975. The method of least squares and some alternatives - Part III, IV, V. *International Statistical Review* 43, 1-44; 125-190; 269-278.
- Hays, W. L. 1973. Statistics for the social sciences. New York: Holt, Rinehart & Winston.
- Herrmann, Th. 1973. Persönlichkeitsmerkmale. Bestimmung und Verwendung in der psychologischen Wissenschaft. Stuttgart: Kohlhammer.
- Hoerl, A. E. & Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal Problems. *Technometrics*, 12, 55-67.
- Holzkamp, K. 1972. Kritische Psychologie. Frankfurt am Main: Fischer.
- Horst, P. 1965. Factor analysis of data matrices. New York: Holt, Rinehart & Winston.
- Hotelling, H. 1933. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* 24, 417-441, 498-520.
- Humak, K. M. S. 1977. Statistische Methoden der Modellbildung. Berlin: Akademieverlag.
- Jöreskog, K. G. 1966. Testing a simple structure hypothesis in factor analysis. *Psychometrika* 31, No. 2, 165-178.
- Jöreskog, K. G. 1967. Some contributions to maximum likelihood factor analysis. *Psychometrika* 32, 443-482.
- Jöreskog, K. G. 1969. A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika* 34, No. 2, 183-202.
- Jöreskog, K. G. 1971. Statistical analysis of sets of congeneric tests. *Psychometrika* 40, 395-412.
- Jöreskog, K. G. 1973. A general method for estimating a linear structural equation System. In: Goldberger, A. S. & Duncan, O. D. (Eds): Structural equation models in the social sciences. New York: Seminar-Press.
- Jöreskog, K. G. & Sörbom, D. 1978. LISREL IV. A general Computer Program for estimation of linear structural equation Systems by maximum likelihood methods. University of Uppsala: Departement of Statistics.

- Kerlinger, F. N. & Pedhazur, E. J. 1973. Multiple regression in behavioral research. New York: Holt, Rinehart & Winston.
- Krauth, J. & Lienert, G. A. 1973. Die Konfigurationsfrequenzanalyse. München: Alber.
- Kreil, C. 1831. Sammlung der notwendigsten math. Formeln aus der Algebra, Trigonometrie, Astronomie und Mechanik. Wien: J. B. Wallishausser.
- Kruskal, W. H. & Tanur, J. M. 1978. International Encyclopedia of Statistics. New York: Free Press.
- Lazarsfeld, P. F. 1950. Logical and mathematical foundations of latent structure analysis. In: Stouffer, S. A. et. al.: Studies in social psychology in world war II, Vol. IV, Princeton N. J. Princeton University Press.
- Lazarsfeld, P. F. 1959. Latent structure analysis. In: Koch, S. (Ed.): Psychologie: A study of a science. Vol. 3, New York: McGraw Hill.
- Lazarsfeld, P. F. & Henry, N. W. 1968: Latent structure analysis. Boston, Houghton Mifflin.
- Lawley, D. N. & Maxwell, A. E. 1971. Factor analysis as a statistical method. London: Butterworths.
- Lienert, G. A. 1969<sup>3</sup>. Testaufbau und Testanalyse. Weinheim: Beltz.
- Littrow, J. J. 1818. Über die gerade Aufsteigung der vornehmsten Fixsterne. Astronomie 6, 1-26.
- Littrow, J. J. 1833. Die Wahrscheinlichkeitsrechnung in ihrer Anwendung auf das wissenschaftliche und praktische Leben. Wien: Becks Universitäts-Buchhandlung.
- Lord, F. M. & Novick, M. R. 1968. Statistical theories of mental test scores. Reading, Mass.: Addison & Wesley.
- Mac Corquodale, K. & Meehl, P. E. 1948. On a distinction between hypothetical constructs and intervening variables. Psychological Review, 95-107.
- Madansky, A. 1978. Latent structure. In: Kruskal, W. A. & Tanur, J. M. (Eds): International Encyclopedia of Statistics. New York: The Free Press.
- Marinell, G. 1977. Multivariate Verfahren. Eine Einführung für Studierende und Praktiker. 1. Aufl. München: Oldenbourg.
- McDonald, R. P. 1967. Nonlinear factor analysis. Psychometric Monograph 15.
- Meehl, P. E. 1950. Configural scoring. Journal of Consulting Psychology. 14, 165-171.
- Moosbrugger, H. 1978. Multivariate statistische Analyseverfahren. Stuttgart: Kohlhammer.
- Moosbrugger, H. 1980. Lineare statistische Modelle zur Beschreibung linearer und nichtlinearer multipler Variablenzusammenhänge. Arbeiten aus dem Institut für Psychologie der JWG-Universität Frankfurt am Main 1980/1.
- Moosbrugger, H. 1981a. Zur differentiellen Validität bei nichtlinearen Test-Kriterium-

- Zusammenhängen. Zeitschrift für Differentielle und Diagnostische Psychologie, 2, 219-234.
- Moosbrugger, H. 1981b. Testtheorien und Testkonstruktion. Arbeiten aus dem Institut für Psychologie der Johann-Wolfgang-Goethe-Universität Frankfurt am Main 1981/6.
- Moosbrugger, H. 1982. Dimensionalitätsuntersuchungen von FPI-Skalen mit dem KLA-Modell. Zeitschrift für Differentielle und Diagnostische Psychologie 3, Heft 4 (im Druck).
- Moosbrugger, H. & Müller, H. 1980. Das klassische latent-additive Testmodell (KLA-Modell) und seine Beziehungen zur klassischen Testtheorie sowie zum dichotomen logistischen Testmodell von Rasch. Wissenschaftliches Wandposter, ausgestellt anlässlich der Poster-Session am 32. Kongreß der Deutschen Gesellschaft für Psychologie in Zürich 1980. Arbeiten aus dem Institut für Psychologie der JWG-Universität Frankfurt am Main 1980/8.
- Moosbrugger, H. & Müller, H. 1981. Ein klassisches latent-additives Testmodell (KLA-Modell). In: W. Michaelis (Ed.): Bericht über den 32. Kongreß der Deutschen Gesellschaft für Psychologie in Zürich 1980. 483-486. Göttingen: Hogrefe.
- Moosbrugger, H. & Müller, H. 1982. A classical latent additive test model (CLA Model). The German Journal of Psychology, 6, 145-149.
- Moosbrugger, H. & Steyer, R. 1982. Uni- und multivariate Varianzanalyse mit festen Parametern. In: Bredenkamp, J. & Feger, H. (Eds): Enzyklopädie der Psychologie, Serie: Forschungsmethoden der Psychologie, Band 4. Göttingen: Hogrefe.
- Müller, H. 1980. über „probabilistische“ und „klassische“ Testmodelle mit latent-additiver Struktur: Neue Aspekte der klassischen Testtheorie. Jahresarbeit. Institut für Psychologie der JWG-Universität Frankfurt am Main.
- Müller, H. & Moosbrugger, H. 1980. Computerprogramm für Parameterschätzung, Item- und Personenanalyse nach dem KLA-Modell. Arbeiten aus dem Institut für Psychologie der JWG-Universität Frankfurt am Main 1980/9.
- Mulaik, S. A. 1972. The foundations of factor analysis. New York: McGraw Hill.
- Orth, B. 1974. Einführung in die Theorie des Messens. Stuttgart: Kohlhammer.
- Pawlak, K. 1968. Dimensionen des Verhaltens. Eine Einführung in Methodik und Ergebnisse faktorenanalytischer psychologischer Forschung. Bern: Huber.
- Pearson, K. 1896. Mathematical contributions to the theory of evolution. III. Regression, Heredity and Panmixia. Philosophical Transactions 187 A, 253-318.
- Pearson, K. 1901. On lines and planes of closest fit to a system of points in space. Philosophical Magazine, 2, 6<sup>th</sup> Series, 557-572.
- Rao, C. R. & Mitra, S. K. 1971. Generalized inverse of matrices and its applications. New York: Wiley.
- Rasch, D. 1960. Probabilistic models for some intelligence and attainment tests. Kopenhagen: The Danish Institute for Educational Research.

- Rasch, G. 1976. Einführung in die mathematische Statistik. Berlin: VEB Deutscher Verlag der Wissenschaften.
- Revenstorf, D. 1976. Lehrbuch der Faktorenanalyse. Stuttgart: Kohlhammer.
- Revenstorf, D. 1980. Faktorenanalyse. Stuttgart: Kohlhammer.
- Rachel, H. 1981. Planung und Auswertung von Untersuchungen im Rahmen des allgemeinen linearen Modells. Psychologisches Institut, Ruhr-Universität Bochum.
- Rost, J. & Spada, H. 1978. Probabilistische Testtheorie. In: K. J. Klauer (Ed.): Handbuch der Pädagogischen Diagnostik, Bd. 1, Düsseldorf: Schwann.
- Ruch, W. & Hehl, F.-J. 1982. Möglichkeiten der Kontrolle der Stichprobenabhängigkeit KLA-skalierter Items bei Vorliegen von Retestdaten. Vortrag auf der 24. Tagung experimentell arbeitender Psychologen in Trier 1982.
- Schermelleh, K. 1979. Vergleichende Darstellung verschiedener Kodierungsverfahren im allgemeinen linearen Modell. Unveröffentlichte Vordiplomarbeit. Institut für Psychologie der JWG-Universität Frankfurt am Main.
- Searle, S. R. 1971. Linear models. New York: Wiley.
- Sörbom, D. & Jöreskog, K. G. 1976. COFAMM: Confirmatory factor analysis with model modifikation. User's Guide. Chicago: National Educational Resources.
- Späth, H. 1973. Algorithmen für elementare Ausgleichsmodelle. München: Oldenbourg.
- Späth, H. 1974. Algorithmen für multivariate Ausgleichsmodelle. München: Oldenbourg.
- Spearman, C. E. 1904. General intelligence objectively determined and measured. American Journal of Psychology, 15, 201-293.
- Spearman, C. E. 1926. The abilities of man. Their nature and measurement. London: Macmillan.
- Steyer, R. 1979. Untersuchungen zur nonorthogonalen Varianzanalyse. Weinheim: Beltz.
- Steyer, R. 1982. Modelle zur kausalen Erklärung statistischer Zusammenhänge. In: Bredenkamp, J. & Feger, H. (Eds): Enzyklopädie der Psychologie, Serie: Forschungsmethoden der Psychologie, Band 4. Göttingen: Hogrefe.
- Steyer, R. & Moosbrugger, H. 1979. Lineare Modelle: Modelle kongenerischer Variablen. Arbeiten aus dem Institut für Psychologie der JWG-Universität Frankfurt am Main 1979/2.
- Steyer, R., Moosbrugger, H. & Ullrich, K. 1981. Rechenregeln für bedingte Erwartungen. Arbeiten aus dem Institut für Psychologie der JWG-Universität Frankfurt am Main 1981/1.
- Tarnai, Ch. 1981. Anwendbarkeit des KLA-Modells bei nichtlinearen Funktionen zwischen latenten und manifesten Reaktionsvariablen. Vortrag auf der 23. Tagung experimentell arbeitender Psychologen in Berlin 1981.

- Tatsuoka, M. M. 1971. Multivariate analysis. Techniques for educational and psychological research. New York: Wiley.
- Thurstone, L. L. 1938. Primary mental abilities. Chicago: The University of Chicago Press.
- Thurstone, L. L. 1947. Multiple factor analysis. Chicago: University of Chicago Press.
- Timm, N. H. 1975. Multivariate analysis with applications in education and psychology. Belmont, Cal.: Wadsworth.
- Überla, K. 1971. Faktorenanalyse. Heidelberg: Springer.
- Weiling, F. 1972. Geschichtliche Notizen zur Entwicklung der Regressionsanalyse. Biometrische Zeitschrift, 14, 164-177.
- Weiling, F. 1973. Die Varianzanalyse. Eine Übersicht mit historischem Aspekt. Vortrag, 19. Kolloquium der Deutschen Region der Internationalen Biometrischen Gesellschaft, Berlin.
- Winer, B. J. 1971<sup>2</sup>. Statistical principles in experimental design. New York: McGraw Hill.
- Wottawa, H. 1974. Das allgemeine lineare Modell. Ein universelles Auswertungsverfahren. EDV in Medizin und Biologie, 3, 65-73.
- Wottawa, H. 1979. Grundlagen und Probleme von Dimensionen in der Psychologie. Meisenheim am Glan: Hain.
- Wottawa, H. 1980. Grundriß der Testtheorie. München: Juventa.
- Yule, G. U. 1897. On the significance of Bravais' formulae for regression etc in the case of skew correlation. Proceedings of the Royal Society London, 60, 477-489.
- Zimmerman, D. W. 1975. Probability spaces, hilbert spaces, and the axioms of test theory. Psychologika, 40, 395-412.

## 2. Kapitel

# Modelle zur kausalen Erklärung statistischer Zusammenhänge

*Rolf Steyer*

## 1. Einführung

### 1.1 Zur Bedeutsamkeit kausaler Abhängigkeit

Statistische Modelle werden in der Psychologie und anderen Bio- und Sozialwissenschaften nicht nur zur Beschreibung von Zusammenhängen der betrachteten Phänomene verwendet, sondern wurden von Anfang an auch mit der Intention der kausalen Erklärung dieser Phänomene eingesetzt. So sollte z.B. Spearmans (1904, 1927) allgemeiner Intelligenzfaktor<sup>1)</sup> die beobachtete Kovariation der verschiedenen Intelligenzleistungen erklären. Wright (1918, 1921, 1923, 1934, 1960a,b) entwickelte seine pfadanalytischen Methoden explizit, um kausale Zusammenhänge genetischer Phänomene zu beschreiben. Fishers (1925) varianzanalytische Verfahren dienten zunächst zur Auswertung landwirtschaftswissenschaftlicher Experimente, bei denen er großen Wert auf Kontrolltechniken wie z.B. Randomisierung gelegt hat (siehe z.B. 1947, S. 17f.), und auch er spricht explizit von „Wirkungen“ (siehe z.B. 1956, S. 241) der experimentell manipulierten Variablen.

In den Forschungsansätzen, die sich um die kausale Erklärung beobachtbarer Zusammenhänge bemühen, nehmen kontrollierte experimentelle und auch quasieperimentelle Untersuchungen (vgl. z.B. Campbell & Stanley, 1963, Cook & Campbell, 1976, 1979) einen bevorzugten Platz ein. Ein großer Teil der Techniken zur Versuchsplanung (siehe z.B. Bartenwerfer & Raatz, 1979, Henning & Muthig, 1979, McGuigan, 1978, Preiser, 1977, Schulz, Muthig & Koeppeler, 1981, Zimmermann, 1972) zielt darauf ab, „interne Validität“ herzustellen, damit man tatsächlich den „Einfluß“ (Bredenkamp, 1969, S. 333) der unabhängigen auf die abhängigen Variablen feststellen kann.

---

<sup>1)</sup> Einen kurzen Überblick über Spearmans Intelligenztheorie, und auch spätere Entwicklungen gibt z.B. Süllwold (1977).

Aber nicht nur in experimentellen Untersuchungen ist die Idee der kausalen Abhängigkeit von grundlegender Bedeutung. Auch in der Theorie latenter Variablen (vgl. z.B. Anderson, 1954, 1959, Gibson, 1959, 1962, Goodman, 1974, 1978, 1979, Green, 1951, 1952, Harnerle, 1982, Jöreskog, 1969, 1971, Lawley & Maxwell, 1971, Lazarsfeld & Henry, 1968, McDonald, 1962, 1967a,b, Moosbrugger, 1982, in diesem Band), mit deren Hilfe man ebenfalls Zusammenhänge beobachtbarer Variablen erklären will, ist die „Idee der Verursachung“ (Lazarsfeld, 1959, S. 500) zentral. Dort will man wissen, „ob die zugrundeliegende Eigenschaft die Beziehungen zwischen den manifesten Indikatoren erklärt“ (Lazarsfeld, 1959, S. 501, Übersetzung durch den Autor). Formann (1980, S. 107) schreibt dazu: „Manifeste Variablen kovariieren nicht deshalb, weil sie an sich miteinander zusammenhängen, sondern weil sie ihre gemeinsame Ursache in den latenten Variablen haben, welche es zu identifizieren gilt.“

Auch die Strukturgleichungsmodelle (vgl. z.B. Duncan, 1975, Goldberger & Duncan, 1973, Jöreskog, 1970, 1973, 1974, Johnston, 1971, Malinvaud, 1970) werden explizit zur Beschreibung kausaler Zusammenhänge verwendet, auch solcher mit wechselseitiger Abhängigkeit. Diese Modelle haben insbesondere durch neuere Arbeiten (siehe z.B. Bentler, 1980, Jöreskog, 1977, 1978, 1979, Jöreskog & Sörbom, 1976, 1977, 1979, 1981, Kenny, 1979) in letzter Zeit große Beachtung gefunden, die es u.a. ermöglichen, kausale Beziehungen auch zwischen latenten Variablen anzunehmen und solche Modelle empirisch zu überprüfen.

Mit Modellen zur Erklärung statistischer Zusammenhänge versucht man, Wirkungsgefüge zu beschreiben, also kausale Theorien zu formulieren. Würde die Psychologie es bei der Beschreibung rein statistischer Zusammenhänge bewenden lassen, wäre sie bestenfalls für Prognosen zu gebrauchen. Darüber hinaus sind jedoch in der Praxis oft Interventionen gefordert, deren Auswirkungen nur vorhersagbar sind, wenn kausale Theorien über das Zusammenwirken der Variablen vorliegen. Brandstädter und Bernitzke (1976, S. 13) plädieren ebenfalls für eine verstärkte Forschung mit kausalen Modellen, indem sie schreiben: „Psychologische Interventionen sind stets Eingriffe in „Systeme“, Eingriffe in komplexe Wirkungsgefüge also, die erwartete und unerwartete, erwünschte und unerwünschte Aus- und Nebenwirkungen haben können. Eine effiziente und risikobewußte Interventionsplanung setzt eine möglichst umfassende Kenntnis des Wirkungsgefüges voraus, in das eingegriffen wird.“

Diese Wirkungsgefüge müssen keineswegs deterministischer Natur sein, vielmehr muß man wohl in der Psychologie und selbst in vielen Bereichen der Physik von stochastischen Gesetzmäßigkeiten ausgehen, worauf schon Sarris (1968) eindringlich hingewiesen hat. Dabei kann durchaus offenbleiben, ob wir zu stochastischen Gesetzen greifen, weil unser Wissen unvollständig ist, in

dem Sinne, daß wir nicht alle beeinflussenden Variablen kennen, oder ob die untersuchten Phänomene an sich nichtdeterministisch sind.

## 1.2 Zum Forschungsstand

In krassem Gegensatz zu der Bedeutsamkeit des Konzepts der kausalen Abhängigkeit stand bisher allerdings die Verschwommenheit, die diesem wohl zentralen Begriff empirischer Forschung eigen ist. Dies liegt keineswegs daran, daß dem Thema ‚kausale Abhängigkeit‘ nicht genügend Aufmerksamkeit geschenkt wird. Vielmehr sind gerade die neueren Versuche, zu einer Klärung dieses Konzepts beizutragen, so zahlreich, daß es unmöglich erscheint, diese auch nur annähernd adäquat darzustellen. Versuche einer solchen Darstellung haben z.B. Bagozzi (1980), Cook & Campbell (1979), Stegmüller (1969) und Suppes (1970) unternommen. In dieser Arbeit beschränken wir uns darauf, den Diskussionsstand in der Psychologie zu umreißen.

In der psychologischen Forschung können unter dem Aspekt kausaler Abhängigkeit wohl zwei Hauptströmungen unterschieden werden, die man vielleicht am besten mit den Etiketten ‚experimentalpsychologische‘ vs. ‚pfadanalytische Literatur‘ versehen kann. Der Experimentalpsychologe bemüht sich, alle ‚Störvariablen‘ auszuschalten oder ‚interne Validität‘ herzustellen, um den Einfluß der unabhängigen auf die abhängigen Variablen feststellen zu können. Dabei vermeiden manche Autoren, von kausaler Abhängigkeit zu sprechen. So z.B. Bredenkamp (1980, S. 1), wenn er formuliert: „Da der Begriff der Verursachung in der wissenschaftstheoretischen Literatur (vgl. Stegmüller, 1969) jedoch anders als in diesem Zusammenhang gebraucht wird, soll das Ziel des Experimentierens darin gesehen werden, die Variation der abhängigen Variablen als Folge der Veränderung von unabhängigen Variablen eindeutig interpretieren zu können.“

Offensichtlich werden hier nur andere Worte für Ursache und kausale Abhängigkeit eingesetzt, die jedoch keinerlei Zweifel daran aufkommen lassen, daß auch hier Ziel des Experimentierens ist, Ursache-Wirkungs-Beziehungen zu untersuchen, auch wenn der dabei verwendete Kausalbegriff nicht weiter expliziert und mit anderen Bezeichnungen umschrieben wird.

Andere Autoren sprechen in diesem Kontext explizit von kausaler Abhängigkeit und verstehen interne Validität als Bedingung, unter der eine kausale Interpretation der gefundenen Abhängigkeiten möglich ist. So schreiben z.B. Cook & Campbell (1979, S. 58): „Interne Validität ist, nach allem, mit den Gefahren befaßt, die Zweifel daran erwecken können, ob ein gültiger kausaler Zusammenhang besteht, . . .“ (Übersetzung durch den Autor).

Genau betrachtet, wird jedoch auch in dieser und ähnlichen Textstellen nur der eine undefinierte Begriff durch einen anderen ebenso undefinierten, nämlich den der ‚internen Validität‘ ersetzt, und der Einfluß einer solchen Methodologie wäre wohl gering geblieben, wenn nicht in der experimentellen Psychologie und Pädagogik zugleich eine Reihe von Handlungsanweisungen entwickelt worden wären, die dazu beitragen, in einer empirischen Untersuchung ‚interne Validität‘ herzustellen. Eine solche Handlungsanweisung ist z.B. die Kontrolltechnik der Randomisierung, bei der die Untersuchungseinheiten (z.B. Versuchspersonen) zufällig auf die experimentellen Gruppen aufgeteilt werden, so daß vor der experimentellen Behandlung keine systematischen Unterschiede zwischen den experimentellen Gruppen bestehen können, und es daher plausibel erscheint, die später gefundenen Unterschiede auf die inzwischen durchgeführte experimentelle Behandlung zurückzuführen. Weitere solche Kontrolltechniken sind z.B. Konstanthaltung von potentiellen Störvariablen und Parallelisierung (vgl. z.B. Selg & Bauer, 1976, S. 58ff.).

Obwohl diese Forschungsstrategien ohne Zweifel von großer Bedeutung sind, bleibt jedoch ungeklärt, ob sie hinreichend für eine kausale Interpretation von Abhängigkeiten sind, die unter solchen kontrollierten Bedingungen gefunden werden. Eine solche Klärung ist nur zu erhoffen, wenn eine von diesen Kontrolltechniken zunächst unabhängige Definition kausaler Abhängigkeit vorliegt, die dem Forschungsgegenstand der Psychologie und anderen Bio- und Sozialwissenschaften angemessen ist.

Die zweite Richtung in der psychologischen Literatur wurde mit dem Etikett ‚Pfadanalyse‘ versehen. Vertreter dieser Methode sind der Auffassung, daß eine kausale Interpretation stochastischer Abhängigkeiten auch in nichtexperimentellen Forschungssituationen möglich ist, wenn man die Annahme der Geschlossenheit machen kann. So schreiben z.B. Namboodiri, Carter und Blalock (1975, S. 446): „. . . , das theoretische System muß als theoretisch geschlossen oder komplett angenommen werden, bis auf strikt zufällige Fehlerquellen in den Gleichungen.“ (Übersetzung durch den Autor). Aber auch hier stellt sich die Frage, was man denn unter „geschlossen“ verstehen soll, und was sind „strikt zufällige Fehlerquellen“? Offenbar ist gemeint, daß alle wichtigen Variablen in der betreffenden Gleichung vorkommen. Aber was heißt dann ‚wichtig‘? Ist z. B. die abhängige Variable gemessen vor einer experimentellen Behandlung - denn pfadanalytische Modelle sollten in experimentellen Situationen wohl erst recht gelten - eine wichtige Variable, da sie vermutlich eine hohe Korrelation mit der abhängigen Variablen gemessen nach der experimentellen Behandlung aufweist, oder ist sie vielleicht nur wichtig, wenn keine der Kontrolltechniken verwendet wurde, um ihren systematischen Einfluß auszuschalten? Wie kann gegebenenfalls die Annahme der Geschlossenheit falsifiziert werden?

Obwohl die Theorien der pfadanalytischen Modelle weit entwickelt sind, was Schätz- und Testmethoden betrifft (siehe z.B. Goldberger, 1964, Heise, 1975, Jöreskog & Sörbom, 1981, Lohmöller, 1981, Wold, 1981), ist in dieser Literatur keine befriedigende Definition kausaler Abhängigkeit zu finden, obwohl nicht wenige Versuche unternommen wurden (vgl. z.B. Simon, 1952, 1953, Wold, 1954, 1956, 1964, 1969). Einer der klarsten Definitionsorschläge stammt von Blalock (1971b, S. 20): „Wir werden sagen, daß X eine direkte Ursache von Y ist (symbolisiert  $X \rightarrow Y$ ) genau dann, wenn wir eine Veränderung im Mittel von Y hervorbringen können, indem wir X verändern, während alle anderen Variablen konstant gehalten werden, die explizit in das System einbezogen sind und die nicht von Y kausal abhängig sind.“ (Übersetzung durch den Autor).

Die m.E. wesentliche Schwäche dieser Definition liegt darin, daß sie sich, genau wie Simons Versuche, nur auf ein willkürliches und zu enges Modell beziehen „- ein Gleichungssystem -, und nicht auf die „reale“ Welt, welche das Modell beschreiben soll“ (Simon, 1953, S. 51, Übersetzung durch den Autor). Bezieht man z.B. nur die Variablen X und Y in das Gleichungssystem ein, so wäre Y - Blalocks Definition zufolge - immer von X kausal abhängig, da ja alle anderen Variablen (hier beträgt deren Anzahl Null) konstant gehalten werden. Folglich wäre Blalocks Aussage nur dann sinnvoll, wenn man die Geschlossenheitsannahme macht, daß nämlich alle wichtigen Variablen im Gleichungssystem enthalten sind, eine Annahme, auf deren Unklarheit bereits hingewiesen wurde.

Dennoch enthält Blalocks Definition wohl bereits die entscheidende Grundidee, wenn man die Konstanzbedingung nicht auf die explizit einbezogenen Variablen beschränkt, sondern dabei alle potentiellen Störvariablen einbezieht, die in der betreffenden Situationsklasse explizit und implizit vorhanden sind. Ein sinnvolles Modell muß also in irgendeiner Form alle potentiellen Störvariablen enthalten. Diese Grundidee und die daraus folgenden Konsequenzen sollen in dieser Arbeit expliziert und präzisiert werden. Dabei stellt sich heraus, daß man unter bestimmten Voraussetzungen auch dann solche kausalen Abhängigkeiten feststellen kann, wenn nicht alle potentiellen Störvariablen konstant sind.

Was Blalock und auch Simon offenbar übersehen, ist vor allem, daß unser Modell nicht nur aus einem Gleichungssystem besteht, im einfachsten Fall also nicht nur aus einer Gleichung, welche die Beziehung zwischen zwei Variablen X und Y angibt. Vielmehr handelt es sich bei unserem Modell um eine formale Struktur, welche die gesamte Untersuchungssituationsklasse oder das gesamte Experiment mit allen darin vorkommenden Ereignissen beschreibt. Diese formale Struktur wird mit  $(\Omega, \mathcal{A}, P)$  bezeichnet und heißt ‚Wahrscheinlichkeitsraum‘. Dabei ist  $\Omega$  die Menge der in der betreffenden Untersuchungssituation vorkommenden Elementareignisse,  $\mathcal{A}$  eine Sigmaalgebra daraus bildbarer

einfacher und zusammengesetzter Ereignisse, und  $P$  ist eine Funktion, die jedem dieser Ereignisse  $A \in \mathcal{A}$  eine feste Wahrscheinlichkeit zuordnet (siehe z.B. Bauer, 1974, S. 129).

Indem jedem einfachen und zusammengesetzten Ereignis eine feste Wahrscheinlichkeit zugeordnet ist, ist eine bestimmte Untersuchungssituationsklasse oder ein Experiment charakterisiert. Dabei müssen diese Wahrscheinlichkeiten keineswegs bekannt sein. Vielmehr ist es u.a. ein Ziel empirischer Untersuchungen, zu datengestützten Schätzungen dieser Wahrscheinlichkeiten zu gelangen.

Wenn nun auf einem solchen Wahrscheinlichkeitsraum, der eine ganz bestimmte Untersuchungssituationsklasse repräsentiert, zwei stochastische Variablen  $X$  und  $Y$  definiert werden, so haben diese beiden Variablen eine ganz bestimmte Wahrscheinlichkeitsverteilung, die möglicherweise auch durch eine Gleichung<sup>2)</sup> wie z.B.

$$(1.2.1) \quad Y \underset{fs}{=} \alpha_{Y0} + \alpha_{YX} X + F, \quad \text{mit } F := Y - E(Y | X),$$

beschrieben werden kann, wobei  $E(Y | X)$  die bedingte Erwartung von  $Y$  unter  $X$ , sowie  $\alpha_{Y0}$  und  $\alpha_{YX}$  reelle Zahlen sind. Unser Modell, und das ist das Entscheidende, besteht aber nicht nur aus dieser Gleichung, wie viele Bücher der angewandten Statistik dem Leser nahelegen; vielmehr schließt dieses Modell den gesamten Wahrscheinlichkeitsraum mit allen darin enthaltenen Ereignissen und deren Wahrscheinlichkeiten ein, und damit auch alle Variablen, die sich auf diesem Wahrscheinlichkeitsraum, in dieser Untersuchungssituationsklasse definieren lassen. In unserem gegenwärtigen Kontext sind dabei vor allem diejenigen stochastischen Variablen  $W$  von Bedeutung, die der Experimentalpsychologe als ‚Störvariablen‘ bezeichnet, und die er mit verschiedenen experimentellen Techniken zu kontrollieren bestrebt ist. So versucht man mit der Kontrolltechnik der Randomisierung zu erreichen, daß die unabhängige Variable  $X$  selbst stochastisch unabhängig von allen potentiellen Störvariablen ist. Dies dient „u.a. dem Zweck, den Einfluß anderer Faktoren, die diesen Zusammenhang (gemeint ist der zwischen  $X$  und  $Y$ , der Autor) modifizieren können, auszuschalten“ (Bredenkamp, 1980, S. 1).

Das implizite Modell des Experimentalpsychologen besteht demnach nicht nur aus einer Gleichung, wie z.B. Gleichung 1.2.1, sondern aus einer Untersuchungssituationsklasse, in der viele, auch unbekannte Variablen eine Rolle

<sup>2)</sup> Das Zeichen „=“ in Gleichung 1.2.1 bedeutet ‚fast sicher gleich‘ oder mit ‚Wahrscheinlichkeit Eins gleich‘, und „ $\underset{fs}{=}$ “ heißt ‚definitionsgemäß fast sicher gleich‘. Die bedingte Erwartung  $E(Y | X)$  ist in der Stochastik so definiert, daß Aussagen über sie nur ‚mit Wahrscheinlichkeit Eins‘ oder ‚fast sicher‘ gelten. Somit werden bei solchen Aussagen ‚Ausnahmen‘ zugelassen, falls diese die Wahrscheinlichkeit Null haben (siehe Definition A.5.1 im Anhang dieser Arbeit).

spielen, und deren Existenz durch eine sorgfältige Wahl und Anwendung verschiedener Kontrolltechniken berücksichtigt wird. Diese verschiedenen Variablen werden nicht einfach ignoriert, sondern die Untersuchungssituationsklasse, das Experiment, wird nach Möglichkeit so gestaltet, daß die anderen Variablen nicht mehr den Zusammenhang zwischen X und Y modifizieren oder verfälschen können. Im Gegensatz zu Blalocks Vorstellungen spielen im Modell des Experimentalpsychologen also nicht nur die explizit genannten Variablen, im einfachsten Fall X und Y, eine Rolle, sondern die gesamte Untersuchungssituationsklasse mit allen darin beobachtbaren Variablen, auch wenn nicht alle diese Variablen tatsächlich beobachtet werden müssen. Mit anderen Worten ausgedrückt, das Modell des Experimentalpsychologen umfaßt den gesamten Wahrscheinlichkeitsraum, der das gesamte Experiment bzw. die gesamte Untersuchungssituationsklasse repräsentiert, mit allen dabei involvierten Ereignissen und Variablen. Die Gleichung zwischen zwei Variablen X und Y ist nur ein einziger von mehreren Bestandteilen seines Modells.

Wenn ein solches Modell unsere Vorstellungen von dem beschreiben soll, was wir als ‚Realität‘ bezeichnen, muß ‚kausal‘ nicht nur eine Eigenschaft des Modells, sondern auch als Eigenschaft dieser ‚Realität‘ postuliert werden; ansonsten wäre eine solche Modelleigenschaft wohl sinnlos. Demnach muß aber auch eine formale Eigenschaft definiert werden, die ein Modell zu einem kausalen Modell macht, und in Anwendungen muß man wenigstens im Sinne von Poppers Falsifizierungskonzept (siehe z.B. Popper, 1969) überprüfen können, ob in dem betreffenden Fall diese Eigenschaft in der ‚Realität‘ erfüllt ist.

Diesen grundlegenden Forderungen entsprechen die bisherigen pfadanalytischen Arbeiten nicht. Die dort üblichen Modelltests prüfen, ob das Modell mit den Daten vereinbar ist und vergleichen verschiedene alternative Modelle hinsichtlich dieses Kriteriums. Sie testen jedoch nicht, ob ein solches Modell kausal interpretierbar ist. Es wird auch nicht angegeben, wie die Annahme der Geschlossenheit falsifiziert werden kann.

Solche Mängel können wohl als mitverantwortlich für viele skeptische Stimmen gegenüber ‚Kausalanalysen‘ und dem Wort ‚kausal‘ überhaupt angesehen werden. Diese Stimmen reichen von Vorschlägen zur Umbenennung bis zur scheinbaren Ablehnung. Sarris z.B. schlägt in seiner Antwort auf Kraaks (1966) Plädoyer für eine verstärkte Kausalforschung die Bezeichnung „statistisch-gesetzmäßige Dependenzen“ vor (Sarris, 1968, S. 184), womit er aber ebenfalls einen stochastischen Kausalbegriff meint. Herrmann (1969, S. 65) scheint einen völlig ablehnenden Standpunkt einzunehmen, wobei er dann aber die ‚neue‘ ebenso diffuse Bezeichnung des „Bedingens“ (ebenda) heranzieht, die wohl auch eher als Umbenennung zu verstehen ist, denn daß mit experimentellen Untersuchungen eben nicht nur rein statistische Zusammenhänge gefunden werden sollen, sondern qualitativ mehr als dies, bestreitet wohl gegenwärtig niemand. Es stellt sich lediglich die Frage, wie diese andere

‘Qualität’, die hier als kausale stochastische Abhängigkeit bezeichnet wird, zu definieren ist. Erst dann lassen sich experimentelle und nichtexperimentelle Forschungsstrategien auf einer soliden theoretischen Grundlage auf Vor- und Nachteile hin diskutieren, und in einer gemeinsamen Theorie integrieren.

### 1.3 Überblick

Im vorliegenden Beitrag wird der Versuch unternommen, die oben aufgeworfenen Probleme durch die Entwicklung einer formalen stochastischen Theorie zu lösen, die zur Grundlegung sowohl experimenteller als auch nichtexperimenteller Strategien der empirischen Kausalforschung herangezogen werden kann. Dabei beschränken wir uns auf regressiv lineare oder kurz ‚reglineare‘ Modelle. Damit ist zwar eine Einschränkung auf relativ einfache Modelle verbunden, aber dennoch fallen unter diese Klasse eine recht große Zahl verschiedener Spezialfälle. Neben den Regressionsmodellen im engeren Sinn, gehören auch Varianz- und faktorenanalytische Modelle zur Klasse der reglinearen Modelle, wie Moosbrugger (1982, in diesem Band) zeigt.

Die grundlegende Frage ist die nach dem Unterschied zwischen einem rein deskriptiven und einem explikativen, kausalen Modell, der nicht nur in der Intention des Anwenders, sondern auch in den formalen Eigenschaften des Modells seinen Niederschlag findet. Aufbauend auf den in der Experimentellen Psychologie und Pädagogik entwickelten Vorstellungen ‚interner Validität‘, werden dabei die Bedingungen der Vorgeordnetheit und der Invarianz als die entscheidenden formalen Eigenschaften entwickelt, die uns ermöglichen, zwischen deskriptiven und explikativen reglinearen Abhängigkeiten aufgrund rein formaler Kriterien zu unterscheiden.

Die Grundidee interner Validität ist die einer idealen Situation, in der sich reine, unkonfundierte stochastische Abhängigkeiten zwischen den betrachteten Ereignissen oder Variablen zeigen. In einer solchen idealen Situation ist der Einfluß anderer Faktoren, die diese Abhängigkeiten modifizieren oder verfälschen können, ausgeschaltet. Dies bedeutet jedoch nicht etwa, daß alle beeinflussenden Faktoren bekannt oder konstant sein müssen. Vielmehr sind lediglich die „konfundierenden Faktoren“ (Sarris, 1968, S. 183) für die interne Validität von Bedeutung, da sie die Beziehung zwischen den betrachteten Variablen verfälschen können. Andere beeinflussende Faktoren sind nur insofern wichtig, als sie die Beziehung zwischen den betrachteten Variablen mit Zufallsfehlern verschleiern, so daß wir die gesuchten Beziehungen mit unseren inferenzstatistischen Mitteln weniger leicht entdecken können.

Die abstrakte Vorstellung einer in diesem Sinn ‚validen‘ oder ‚unkonfundierten‘ Beziehung zwischen zwei stochastischen Variablen (Zufallsvariablen) kann formal, d.h. ohne jeden Rekurs auf formal undefinierte Begriffe wie

,Experiment‘, „Zeit‘, „Manipulation‘ etc. formuliert werden (vgl. dagegen Wold, 1969, S. 448 f.). Der Präzisierung dieser Vorstellung zum Begriff einfacher kausaler reglinearer Abhängigkeit und den Folgerungen daraus sind die Abschnitte 3 bis 5 gewidmet, nachdem wir uns im zweiten Abschnitt anhand eines anschaulichen Beispiels mit konkretem Anschauungsmaterial ausgerüstet haben. In Abschnitt 6 wird dann wieder ein Beispiel behandelt, an dem dann in Abschnitt 7 Fragen des Geltungsbereichs oder der „externen Validität“ (siehe z.B. Campbell & Stanley, 1963) diskutiert werden können. Im letzten Abschnitt wird schließlich in einer Schlußbetrachtung noch einmal die hier vorgestellte Theorie in Hinsicht auf die sich aus ihr ergebenden praktischen Konsequenzen für die empirische experimentelle und nichtexperimentelle Kausalforschung zusammengefaßt.

Obwohl es in diesem Beitrag nur um die Abhängigkeit zwischen zwei stochastischen Variablen geht, sind die hier entwickelten Ideen auch für komplexere Modelle von Bedeutung. Indem man jeweils eine abhängige und eine unabhängige Variable bei Konstanz der anderen unabhängigen Variablen betrachtet, kann die Theorie einfacher kausaler reglinearer Abhängigkeit auch für Mehrvariablenprobleme zur Theorie direkter und totaler kausaler reglinearer Abhängigkeit fortgeführt werden (siehe Steyer, 1982).

Einige weitere Worte zu dem, was in der vorliegenden Arbeit nicht behandelt wird, scheinen angebracht: In der gesamten Abhandlung werden keinerlei Stichprobenprobleme wie z.B. Hypothesen- und Modellbewertung behandelt. Statt dessen untersuchen wir alle hier diskutierten Fragen gewissermaßen auf „Populationsebene“, d.h. wir verfahren so, als wären alle Parameter bekannt und müßten nicht erst aus einer Stichprobe geschätzt werden. Deswegen entstehen auch keinerlei statistische Entscheidungsprobleme zwischen verschiedenen Hypothesen über die Parameter, Probleme, die in Anwendungsfällen natürlich in der Regel vorhanden sind. Für solche Stichprobenprobleme sei daher außer auf die bereits genannten Arbeiten auf die einschlägigen Standardwerke zur multivariaten Statistik verwiesen, wie z.B. Ahrens (1968), Anderson (1958), Bock (1975), Bortz (1977), Gaensslen und Schubö (1973), Krishnaiah (1980), Moosbrugger (1978), Rao (1973), Rasch (1976, 1978), Schach und Schäfer (1978), Scheffe (1959), Searle (1971), Tatsuoka (1971), Timm (1975) und Winer (1971).

## 2. Münzen und Elektromagnet

### 2.1 Einleitende Bemerkungen

Bevor wir eine formale Behandlung kausaler stochastischer Modelle in Angriff nehmen, soll zunächst an einem einfachen Beispiel eine solche kausale stocha-

stische Abhangigkeit beschrieben werden. Wir verfolgen dabei zwei Ziele. Zum einen wollen wir zeigen, da deterministische Theorien kausaler Abhangigkeit nicht allgemein genug sind, um Abhangigkeiten zu beschreiben, die zwar intuitiv die Bezeichnung ‚kausal‘ verdienen, aber dennoch nur stochastisch sind. Zum anderen soll an diesem Beispiel die formale Darstellung in den spateren Abschnitten veranschaulicht und die zentrale Frage nach dem formalen Unterschied zwischen einem kausalen und einem nichtkausalen Regressionsmodell erlautert werden.

## 2.2 Beschreibung des Beispiels

Das Beispiel besteht aus folgendem Gedankenexperiment: Wir betrachten eine Munze I und eine Munze II, die auf jeweils einer Seite aus Plastik und auf der anderen jeweils aus Metall bestehen. Mit diesen Munzen stellen wir uns einen Versuch vor, der folgendermaen aussieht: Zuerst wird Munze II, danach Munze I geworfen, und zwar auf eine Platte, welche die Eigenschaften eines ein- und ausschaltbaren Elektromagneten besitzt.

Tabelle 2.2.1: Erste Situationsklasse: Der Elektromagnet ist ausgeschaltet, wahrend die beiden Munzen geworfen werden. Die Ergebnisse des Werfens der Munzen I und II sind bedingt stochastisch unabhangig (siehe Gleichung 2.2.4). Bei den angegebenen Zahlen handelt es sich um bedingte Wahrscheinlichkeiten gegeben das Ereignis  $\bar{A}$ , da der Elektromagnet ausgeschaltet ist.

		Munze II fallt auf		
		Metallseite C	Plastikseite $\bar{C}$	
Munze I fallt auf	Metallseite B	0.25	0.25	0.50
	Plastikseite B	0.25	0.25	0.50
		0.50	0.50	1.00

1. Situationsklasse: Der Elektromagnet ist ausgeschaltet. In der ersten Situationsklasse wird der Versuch: ‚Zuerst Munze II, danach Munze I werfen‘ einmal durchgefuhrt, und zwar dann, wenn der Elektromagnet ausgeschaltet ist. Wir nehmen dabei der Einfachheit halber an, da es sich um zwei bezglich ihrer physikalischen Eigenschaften gleiche Munzen handelt, insofern als beide mit einer Wahrscheinlichkeit von 0.5 auf die Metall- bzw. Plastikseite fallen. Wenn der Elektromagnet aus ist, sind unserer intuitiven Theorie zufol-

ge die Ergebnisse des Werfens beider Münzen sowohl stochastisch als auch kausal voneinander unabhängig. Betrachten wir z.B. die Ereignisse

(2.2.1)  $B :=$  Münze I fällt auf die Metallseite,

und

(2.2.2)  $C :=$  Münze II fällt auf die Metallseite,

so erhalten wir demnach die bedingte Wahrscheinlichkeit, daß beide Münzen auf die Metallseite fallen gegeben das Ereignis

(2.2.3)  $\bar{A} :=$  der Elektromagnet ist ausgeschaltet,

über die Gleichung

$$(2.2.4) \quad P(B \cap C | \bar{A}) = P(B | \bar{A}) \cdot P(C | \bar{A}) = 0.5 \cdot 0.5 = 0.25$$

(siehe Tabelle 2.2.1). Die angegebenen Wahrscheinlichkeiten kann man sich als relative Häufigkeiten einer Versuchsreihe vorstellen, bei der die einzelnen Versuche, „zuerst Münze II, danach Münze I werfen“, sehr oft durchgeführt wurden. Nachdem diese Wahrscheinlichkeiten einmal bekannt sind, charakterisieren sie jedoch auch jeden einzelnen Versuch<sup>3)</sup>.

Tabelle 2.2.2: Zweite Situationsklasse: Der Elektromagnet ist angeschaltet, während die beiden Münzen geworfen werden. Die Ergebnisse des Werfens der Münzen I und II sind auch hier bedingt stochastisch unabhängig (siehe Gleichung 2.2.6), lediglich die bedingten Randwahrscheinlichkeiten sind gegenüber Tabelle 2.2.1 verändert (z.B. 0.9 statt 0.5). Die angegebenen Zahlen sind bedingte Wahrscheinlichkeiten gegeben das Ereignis A, daß der Elektromagnet angeschaltet ist.

		Münze II fällt auf		
		Metallseite C	Plastikseite C	
Münze I fällt auf	Metallseite B	0.81	0.09	0.90
	Plastikseite $\bar{B}$	0.09	0.01	0.10
		0.90	0.10	1.00

<sup>3)</sup> Vergleiche zur Interpretation des Wahrscheinlichkeitsbegriffs z.B. Bunge, 1967, S. 423f., Popper, 1959, oder Stegmüller, 1973, S. 129.

2. Situationsklasse: Der Elektromagnet ist angeschaltet. Die zweite Situationsklasse unterscheidet sich von der ersten dadurch, daß der Elektromagnet angeschaltet ist, während die beiden Münzen geworfen werden. Dies hat zur Folge, daß die bedingte Wahrscheinlichkeit gegeben

$$(2.2.5) \quad A := \text{der Elektromagnet ist angeschaltet},$$

für beide Münzen auf die Metallseite zu fallen erhöht ist, und zwar, wie wir annehmen, auf 0.9. Demgemäß ist nun 0.1 die bedingte Wahrscheinlichkeit für beide Münzen, auf die Plastikseite zu fallen. Auch in dieser Situationsklasse sind die Ergebnisse des Werfens beider Münzen sowohl statistisch, als unserer intuitiven Vorstellung zufolge, auch kausal voneinander unabhängig, und entsprechend finden wir auch hier z.B. die Wahrscheinlichkeit dafür, daß beide Münzen auf die Metallseite fallen durch die Multiplikation der einzelnen bedingten Wahrscheinlichkeiten, also

$$(2.2.6) \quad P(B \cap C | A) = P(B | A) \cdot P(C | A) = 0.9 \cdot 0.9 = 0.81$$

(siehe Tabelle 2.2.2).

Tabelle 2.2.3: Dritte Situationsklasse: Der Elektromagnet ist mit Wahrscheinlichkeit 0.5 an- bzw. ausgeschaltet, während die beiden Münzen geworfen werden. Die Ergebnisse des Werfens der Münzen I und II sind nicht stochastisch unabhängig (siehe Ungleichung 2.2.7). Bei den angegebenen Zahlen handelt es sich um unbedingte Wahrscheinlichkeiten.

		Münze II fällt auf		
		Metallseite C	Plastikseite $\bar{C}$	
Münze I fällt auf	Metallseite B	0.53	0.17	0.70
	Plastikseite $\bar{B}$	0.17	0.13	0.30
		0.70	0.30	1.00

3. Situationsklasse: Der Elektromagnet ist mit Wahrscheinlichkeit 0.5 an- oder ausgeschaltet. Die dritte Situationsklasse besteht darin, daß der Versuch ‚Zuerst Münze II, danach Münze I werfen‘ unter der Bedingung durchgeführt wird, daß der Zustand des Elektromagneten unbekannt, und dabei aber mit Wahrscheinlichkeit 0.5 an- oder ausgeschaltet ist, während die beiden Münzen geworfen werden. Die für diese Situationsklasse angegebenen Wahrscheinlichkeiten ergeben sich hier als Mittel der entsprechenden bedingten Wahrscheinlichkeiten in den ersten beiden Situationsklassen.

In dieser dritten Situationsklasse ist ein stochastischer Zusammenhang zwischen den Ergebnissen des Werfens der Münze I und denen des Werfens der Münze II festzustellen. Dabei sind die Münzwurfergebnisse zwar stochastisch (siehe Tabelle 2.2.3), unserer intuitiven Vorstellung zufolge aber nicht kausal voneinander abhängig. Statt dessen sind die Ergebnisse des Werfens der beiden Münzen vom Zustand des Elektromagneten kausal stochastisch abhängig. Ist der Elektromagnet nämlich an, so fallen die beiden Münzen eher auf die Metallseite, und ist er aus, so fallen sie mit unveränderter Wahrscheinlichkeit auf die Metallseite. Variiert der Zustand des Elektromagneten, wie in dieser dritten Situationsklasse, so kommt auf diese Weise auch eine Kovariation der Münzwurfereignisse zustande. Diese stochastische Abhängigkeit äußert sich darin, daß nun anstelle der Gleichungen 2.2.4 oder 2.2.6 die Ungleichung

$$(2.2.7) \quad P(B \cap C) = 0.53 \neq P(B) \cdot P(C) = 0.7 \cdot 0.7 = 0.49$$

gilt.

Das hier beschriebene Beispiel<sup>4)</sup> „Münzen und Elektromagnet“ ist in verschiedener Hinsicht für die in dieser Arbeit behandelte Thematik von Interesse, denn es kommen sowohl nichtkausale als auch kausale stochastische Abhängigkeiten vor, nämlich nichtkausale stochastische Abhängigkeiten zwischen den Münzwurfereignissen, und kausale stochastische Abhängigkeiten jeweils eines Münzwurfereignisses vom Zustand des Elektromagneten. Ist dieser nämlich an, so erhöhen sich die bedingten Wahrscheinlichkeiten, für jede der beiden Münzen auf die Metallseite zu fallen von 0.5 auf 0.9, und offenbar können wir mit dem Elektromagneten diese Wahrscheinlichkeiten steuern, obwohl keine deterministische Abhängigkeit der Münzwurfereignisse vom Zustand des Elektromagneten vorliegt (vgl. zum Handlungsaspekt des Kausalbegriffs auch von Wright, 1974).

Um ein analoges Beispiel aus der Psychologie zu erhalten, brauchen wir lediglich an die Stelle der beiden Münzen zwei Versuchspersonen zu setzen, die z.B. auf einen akustischen Reiz mit „ja, ich höre“ bzw. „nein, ich höre nicht“ antworten. Die Stelle des Elektromagneten nimmt dann der akustische Reiz ein, der in zwei Ausprägungen dargeboten wird.

<sup>4)</sup> Ein von der formalen Struktur analoges Beispiel findet sich in Lazarsfeld und Henry (1968, S. 17). Anstelle der beiden Münzen I und II handelt es sich dort um zwei Zeitschriften A und B, die von den befragten Personen gelesen oder nicht gelesen werden, was dem Fallen der Münzen auf die Metall- bzw. Plastikseite entspricht, und an die Stelle des Elektromagneten, der die stochastischen Abhängigkeiten zwischen den Münzwurfereignissen verursacht, tritt die latente Variable „hoher vs. niedriger Ausbildungsstand“, welche die Kovariation des Lesens der beiden Zeitschriften herbeiführt.

### 2.3 Abhangigkeit der ersten von der zweiten Munzvariablen

Wir formulieren nun die in der dritten Situationsklasse (siehe Tabelle 2.2.3)<sup>5)</sup> festzustellende stochastische Abhangigkeit des Ergebnisses des Werfens der Munze I von dem Ergebnis des zuerst erfolgenden Werfens der Munze II, wobei wir bewußt den Elektromagneten ignorieren. Die Abhangigkeit dieser Munzwurfergebnisse lasst sich als reglineare Abhangigkeit wie folgt beschreiben: Wir nehmen an, da die Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  die dritte Situationsklasse repräsentiert und definieren die beiden Munzvariablen

$$(2.3.1) \quad Z_I := \begin{cases} 1, & \text{wenn die Munze I auf die Metallseite,} \\ 0, & \text{wenn sie auf die Plastikseite fallt,} \end{cases}$$

$$(2.3.2) \quad Z_{II} := \begin{cases} 1, & \text{wenn die Munze II auf die Metallseite,} \\ 0, & \text{wenn sie auf die Plastikseite fallt,} \end{cases}$$

und das Ereignis

$$(2.3.3) \quad B := \{\omega \in \Omega : Z_I(\omega) = 1\},$$

da die Munze I auf die Metallseite fallt. Fur die Beziehung der beiden Munzvariablen gilt dann die Gleichung<sup>6)</sup>

$$(2.3.4) \quad E(Z_I | Z_{II}) = \beta_{10} + \beta_{12} Z_{II},$$

wobei  $\beta_{10}$  und  $\beta_{12}$  reellwertige Konstanten sind. Aus dieser Gleichung folgen, unter Benutzung der Gleichungen A.5.2, A.4.6 und A.4.3 des Anhangs

$$(2.3.5) \quad E(Z_I | Z_{II}=1) = P(B | Z_{II}=1) = \beta_{10} + \beta_{12}$$

und

$$(2.3.6) \quad E(Z_I | Z_{II}=0) = P(B | Z_{II}=0) = \beta_{10} = 0.17/0.3 = 0.5666$$

(siehe Tabelle 2.2.3). Demnach ist also  $\beta_{10}$  bei diesem Beispiel als bedingte Wahrscheinlichkeit fur B zu interpretieren, unter der Bedingung namlich, da  $Z_{II} = 0$  ist. Fur  $\beta_{12}$  gilt dann gemas den Gleichungen 2.3.5 und 2.3.6

$$(2.3.7) \quad \begin{aligned} \beta_{12} &= P(B | Z_{II}=1) - P(B | Z_{II}=0) = \\ &= 0.53/0.7 - 0.5666 = 0.1905 \end{aligned}$$

(siehe wiederum Tabelle 2.2.3). Auf analoge Weise konnen wir auch die Parameter fur die Gleichung

<sup>5)</sup> Bei Referenzen mit drei Ziffern geben die ersten beiden Ziffern immer den Abschnitt an.

<sup>6)</sup> Man beachte die Funote zur Gleichung 1.2.1.

$$(2.3.8) \quad E(Z_{II} | Z_I) \stackrel{fs}{=} \beta_{20} + \beta_{21} Z_I$$

berechnen, welche die stochastische Abhängigkeit der zweiten von der ersten Münzvariablen angibt. Dabei erhalten wir  $\beta_{20} = 0.5666$  und  $\beta_{21} = 0.1905$ .

Die stochastischen Variablen  $Z_I$  und  $Z_{II}$  sind nach dem Gesichtspunkt der Handlungsabfolge geordnet, da zuerst die Münze II und dann die Münze I geworfen wird. Dennoch würde es unserer Intuition widersprechen,  $\beta_{12}$  kausal zu interpretieren, nur weil  $Z_{II}$  der Variablen  $Z_I$  vorgeordnet ist. Statt dessen handelt es sich bei  $\beta_{12}$  um einen Koeffizienten, der eine nichtkausale reglineare Abhängigkeit beschreibt, bei der  $Z_{II}$  der Variablen  $Z_I$  durch die festgelegte Handlungsabfolge (zuerst Münze II, dann Münze I werfen) vorgeordnet ist. Die zeitliche Vorgeordnetheit der Variablen ist demnach keine hinreichende Bedingung dafür, daß eine bestehende stochastische Abhängigkeit auch kausal zu interpretieren ist (vgl. z.B. Zimmermann, 1972, S. 40, aber auch z.B. Granger, 1969, 1973, Hosoya, 1977 und Pierce & Haugh, 1977).

## 2.4 Abhängigkeit der Münz- von der Magnetvariablen

Die in der dritten Situationsklasse bestehende stochastische Abhängigkeit zwischen den beiden Münzvariablen läßt sich kausal durch die Variation des Zustands des Elektromagneten erklären. Offenbar handelt es sich um eine kausale Abhängigkeit, die stochastischer Art ist, womit allerdings nicht in Frage gestellt werden soll, daß es prinzipiell möglich ist, die Münzwurfergebnisse deterministisch zu erklären. Allerdings wäre eine solche Theorie weitaus komplexer und schwerer anzuwenden, als die hier behandelte stochastische Theorie, da nicht nur der Zustand des Elektromagneten bekannt sein müßte, sondern auch die Drehgeschwindigkeit der Münzen in allen drei Dimensionen des Raumes, die Wurfhöhe etc.

Für die hier zu besprechende stochastische Theorie benötigen wir lediglich noch die Magnetvariable

$$(2.4.1) \quad Z_M := \begin{cases} 1, & \text{wenn der Elektromagnet an-} \\ & \text{geschaltet ist.} \\ 0, & \text{wenn er ausgeschaltet ist.} \end{cases}$$

Die Abhängigkeiten der Münzvariablen  $Z_I$  von  $Z_{II}$  und der Magnetvariablen  $Z_M$  sowie der Variablen  $Z_{II}$  von  $Z_M$  lassen sich durch die beiden Gleichungen

$$(2.4.2) \quad E(Z_I | Z_{II}, Z_M) \stackrel{fs}{=} P(B | Z_{II}, Z_M) \stackrel{fs}{=} \alpha_{10} + \alpha_{12} Z_{II} + \alpha_{13} Z_M = 0.5 + 0 \cdot Z_{II} + 0.4 \cdot Z_M$$

und

$$(2.4.3) \quad E(Z_{II} | Z_M) \stackrel{fs}{=} P(C | Z_M) \stackrel{fs}{=} \alpha_{20} + \alpha_{23} Z_M = 0.5 + 0.4 \cdot Z_M$$

beschreiben, wobei B das Ereignis ist, daß die Münze I auf die Metallseite fällt, und

$$(2.4.4) \quad C := \{\omega \in \Omega : Z_{II}(\omega) = 1\}$$

das Ereignis ist, daß die Münze II auf die Metallseite fällt.

Die numerischen Werte der in diesen Gleichungen vorkommenden Koeffizienten können aufgrund folgender Überlegungen berechnet werden: Aus den beiden Gleichungen 2.4.2 und 2.4.3 folgen nach den Gleichungen A.5.2, A.4.6 und A.4.3 des Anhangs

$$(2.4.5) \quad P(B | Z_{II}=0, Z_M=0) = \alpha_{10} = 0.5,$$

$$(2.4.6) \quad P(B | Z_{II}=0, Z_M=1) - P(B | Z_{II}=0, Z_M=0) = \alpha_{13} = 0.9 - 0.5 = 0.4,$$

$$(2.4.7) \quad P(B | Z_{II}=1, Z_M=0) - P(B | Z_{II}=0, Z_M=0) = \alpha_{13} = 0.5 - 0.5 = 0,$$

$$(2.4.8) \quad P(C | Z_M=0) = \alpha_{20} = 0.5,$$

und

$$(2.4.9) \quad P(C | Z_M=1) - P(C | Z_M=0) = \alpha_{23} = 0.9 - 0.5 = 0.4,$$

wobei die in den Tabellen 2.2.1 bis 2.2.3 angegebenen Wahrscheinlichkeiten zugrunde liegen.

Aus den Gleichungen 2.4.2 und 2.4.3 lassen sich mit den in den Gleichungen 2.4.5 bis 2.4.9 angegebenen Koeffizienten die stochastischen Abhängigkeiten zwischen  $Z_I$  und  $Z_{II}$  ableiten, und in diesem Sinn erklären, denn sie implizieren die folgende Gleichung für die Kovarianz der Variablen  $Z_I$  und  $Z_{II}$ :

$$(2.4.10) \quad C(Z_I, Z_{II}) = C[(\alpha_{10} + \alpha_{13} Z_M + F_I), (\alpha_{20} + \alpha_{23} Z_M + F_{II})],$$

mit

$$(2.4.11) \quad F_I := Z_I - E(Z_I | Z_M)$$

und

$$(2.4.12) \quad F_{II} := Z_{II} - E(Z_{II} | Z_M).$$

Die Kovarianz dieser beiden Residualvariablen mit  $Z_M$  ist jeweils Null (siehe die Gleichungen A.5.8 und A.5.13):

$$(2.4.13) \quad C(F_I, Z_M) = C(F_{II}, Z_M) = C(F_I, F_{II}) = 0,$$

und das gleiche gilt wegen  $E(F_I | Z_{II}, Z_M) = 0$  und Gleichung A.5.12 auch für die Kovarianz untereinander, was aus den Theoremen A.5.2. und A.5.4 folgt.

Nach den Gleichungen A.3.2, A.3.4 und A.3.5 folgt daher die Kovarianzstrukturgleichung

$$(2.4.14) \quad C(Z_I, Z_{II}) = \alpha_{13} \cdot C(Z_M, Z_M) \cdot \alpha_{23} = \alpha_{13} \cdot V(Z_M) \cdot \alpha_{23} = \\ = 0.4 \cdot 0.25 \cdot 0.4 = 0.04.$$

Dabei haben wir die Varianz  $V(Z_M)$  unter Benutzung von  $E(Z_M) = P(A) = 0.5$  berechnet (siehe die Gleichungen A.3.2 und A.2.3 und Tabelle 2.2.3), wobei A das Ereignis ist, daß der Elektromagnet angeschaltet ist.

Gleichung 2.4.14 entspricht dem sogenannten Grundtheorem der Pfadanalyse (siehe z.B. Seibel & Nygreen, 1972, S. 7 oder Wright, 1934, S. 163). Der einzige Unterschied besteht darin, daß bei Wright nicht die Kovarianzen, sondern die Korrelationen zugrunde gelegt werden. Gleichung 2.4.14 ist jedoch keine besondere Eigenschaft kausaler Modelle, wie man manche Autoren (vgl. z.B. Hummell & Ziegler, 1976b, S. E 61 ff.) verstehen könnte, da sie bereits aus den Gleichungen 2.4.2 und 2.4.3 folgt.

$C(F_I, F_{II}) = 0$  besagt, daß die Erklärung der Kovarianz von  $Z_I$  und  $Z_{II}$  durch die Gleichungen 2.4.2 und 2.4.3 perfekt ist. Dies kann jedoch nicht als Kriterium für die kausale Interpretierbarkeit der durch die Gleichungen 2.4.2 und 2.4.3 beschriebenen Abhängigkeiten herangezogen werden. Würden wir nämlich nur die Münze II werfen, dann würde Gleichung 2.4.3 immer noch die kausale Abhängigkeit der Münzvariablen  $Z_{II}$  von der Magnetvariablen  $Z_M$  beschreiben, obwohl dann gar keine Kovarianz  $C(F_I, F_{II})$  vorkäme.

Die auf die oben beschriebene Weise ohne zusätzliche Annahmen aus den Gleichungen 2.4.2 und 2.4.3 abgeleitete Kovarianz zwischen  $Z_I$  und  $Z_{II}$  stimmt exakt mit derjenigen Kovarianz für diese beiden Variablen überein, die wir direkt aus den Daten der Tabelle 2.2.3 berechnen können.

Die Kovarianz zwischen zwei stochastischen Variablen X und Y mit den Erwartungswerten  $E(X)$  und  $E(Y)$  ist definiert durch

$$(2.4.15) \quad C(X, Y) := E([X - E(X)] \cdot [Y - E(Y)]).$$

Daraus und aus der für eine diskrete stochastische Variable Z mit Erwartungswert  $E(Z)$  und den Werten  $z_i$  gültigen Definitionsgleichung A.2.3 des Anhangs für den Erwartungswert berechnen wir nun die Kovarianz  $C(Z_I, Z_{II})$ . Wenden wir die Gleichung A.2.3 auf das in Gleichung 2.4.15 vorkommende Produkt  $[X - E(X)] \cdot [Y - E(Y)]$  an, so erhalten wir

$$(2.4.16) \quad C(Z_I, Z_{II}) = (1 - 0.7) \cdot (1 - 0.7) \cdot P(B \cap C) + \\ + (1 - 0.7) \cdot (0 - 0.7) \cdot P(B \cap \bar{C}) + \\ + (0 - 0.7) \cdot (1 - 0.7) \cdot P(\bar{B} \cap C) + \\ + (0 - 0.7) \cdot (0 - 0.7) \cdot P(\bar{B} \cap \bar{C}) = \\ = 0.09 \cdot 0.53 + (-0.21) \cdot 0.17 + \\ + (-0.21) \cdot 0.17 + 0.49 \cdot 0.13 = \\ = 0.04.$$

Da die beiden Variablen nur zwei Werte annehmen können, ist mit der Kovarianz die gesamte stochastische Abhängigkeit zwischen den beiden Münzvaria-

blen  $Z_I$  und  $Z_{II}$  beschrieben. Sie ist in diesem Sinn durch die in den Gleichungen 2.4.2 und 2.4.3 beschriebene Abhangigkeit dieser beiden Variablen von der Magnetvariablen  $Z_M$  erklart.

## 2.5 Das Problem

Unserer Intuition gemab sind die Koeffizienten  $\alpha_{23}$ ,  $\alpha_{12}$  und  $\alpha_{13}$  aus den Gleichungen 2.4.2 bzw. 2.4.3, welche die Abhangigkeiten der Munz- von der Magnetvariablen beschreiben, kausal zu interpretieren. Es stellt sich nun jedoch die Frage, welches eigentlich die formalen Unterschiede zwischen der hier vorliegenden kausalen Abhangigkeit der Munz- von der Magnetvariablen und der durch Gleichung 2.3.4 beschriebenen nichtkausalen reglinearen Abhangigkeit der Munzvariablen  $Z_I$  von der Munzvariablen  $Z_{II}$  ist. In beiden Fallen liegen korrekte Regressionsgleichungen<sup>7)</sup> vor, aber dennoch beschreibt die eine unserer Intuition nach nur eine nichtkausale, die andere hingegen eine kausale reglineare Abhangigkeit.

Das in der experimentellen psychologischen und padagogischen Forschung hierfur entwickelte Kriterium ist das der ‚internen Validitat‘, welches im nachfolgenden Abschnitt behandelt werden soll.

## 2.6 Zusammenfassende Bemerkungen

In diesem Kapitel wurde ein Beispiel behandelt, in dem sowohl intuitiv kausale als auch nichtkausale Abhangigkeiten vorkommen, die erstens alle nichtdeterministisch sind, und sich zweitens alle durch eine Regressionsgleichung korrekt beschreiben lassen. Es wurde gezeigt, daß weder diese Beschreibbarkeit durch eine Regressionsgleichung, noch eine zeitliche Geordnetheit der Variablen ein hinreichendes Kriterium zur Unterscheidung der kausalen von der nichtkausalen Abhangigkeit sein kann. Somit ist die Frage aufgeworfen, welche Bedingung als ein solches hinreichendes Kriterium fur eine kausale Interpretierbarkeit stochastischer Abhangigkeiten herangezogen werden kann.

## 3. Interne Validitat

### 3.1 Einleitende Bemerkungen

Der Begriff der internen Validitat wurde neben dem der externen Validitat (siehe Abschnitt 7) als eines von zwei Gutekriterien empirischer Untersuchun-

---

<sup>7)</sup> Die genannten Abhangigkeiten konnten jeweils auch mit einer Strukturgleichung beschrieben werden. In den hier vorliegenden Fallen aber ware eine solche Strukturgleichung mit der entsprechenden Regressionsgleichung formal ´aquivalent.

gen von Donald T. Campbell (siehe z.B. Campbell, 1957 sowie Campbell & Stanley, 1963) eingeführt. Campbell und Stanley (1963, S. 175) schreiben: „Interne Validität ist das grundlegende Minimum, ohne das jedes Experiment uninterpretierbar ist: Haben tatsächlich die experimentellen Behandlungen in diesem Fall einen Unterschied herbeigeführt?“ (Übersetzung durch den Autor).

Noch deutlicher, und in ihrer Formulierung nicht mehr nur auf experimentelle Untersuchungen beschränkt, formulieren Cook und Campbell (1979, S. ix) die Beziehung zwischen interner Validität und kausaler Abhängigkeit, indem sie schreiben: „Als Bedrohung der internen Validität sind Faktoren anzusehen, die zu ungültigen Schlüssen darüber führen können, ob eine Beziehung zwischen manipulierten oder gemessenen Variablen einen kausalen Einfluß von der einen auf die andere widerspiegelt.“ (Übersetzung durch den Autor). Offenbar verstehen Cook und Campbell unter interner Validität eine Bedingung, unter der eine kausale Interpretation stochastischer Abhängigkeit möglich ist. Die Vorstellung, daß kausale Abhängigkeit sich nur unter bestimmten idealen Bedingungen auch in einer entsprechenden stochastischen Abhängigkeit äußert, ist wohl recht plausibel, denn auch deterministische physikalische Gesetze gelten ja nur unter idealen Voraussetzungen, in denen störende Einflüsse ausgeschaltet sind. Man denke nur an das Laub, das, obwohl dem Fallgesetz unterworfen, nicht fällt, sondern in einem warmen Luftstrom, dem Fallgesetz und zugleich den Gesetzen der Aerodynamik gehorcheinnd nach oben getragen wird.

In diesem Abschnitt ist es unser Ziel, die intuitiven Vorstellungen, die dem Begriff der internen Validität zugrunde liegen, an einigen Beispielen zu erläutern, auf einige Unterschiede unseres Verständnisses zu den von anderen Autoren vertretenen Versionen dieses Konzepts hinzuweisen und damit die im nächsten Kapitel beginnende formale Behandlung vorzubereiten.

### 3.2 Grundideen

Die Grundidee einer intern validen Abhängigkeit einer Variablen Y von einer zweiten Variablen X besteht im wesentlichen aus der Vorstellung, daß keine anderen Variablen W existieren, die den Zusammenhang zwischen X und Y modifizieren (vgl. Bredenkamp, 1980, S. 1). Bei dieser Formulierung wird nicht berücksichtigt, ob W zwischen X und Y angeordnet ist oder nicht. Bedenken wir den Fall, daß eine Variable X eine zweite Variable Y nur indirekt über eine dritte Variable W beeinflußt, so kann durchaus der Fall eintreten, daß bei Konstanthaltung von W der Zusammenhang zwischen X und Y verschwindet. Dies ist jedoch m.E. kein Grund die Validität dieses Zusammenhangs in Frage zu stellen, sondern lediglich seine Direktheit. Ein kausaler

Zusammenhang zwischen zwei Variablen X und Y ist m. E. ohne zwischen X und Y vermittelnde Variablen gar nicht denkbar. Als potentielle Störvariablen, welche die Beziehung zwischen X und Y modifizieren können, sollen daher im folgenden nicht alle Variablen betrachtet werden, sondern nur diejenigen Variablen W, die X gleich- oder vorgeordnet sind.

In diesem Sinn modifiziert, besagt also die hier vertretene Idee einer intern validen Abhängigkeit, daß bei Konstanz einer beliebigen potentiellen Störvariablen W auf einem ihrer Werte w, die Abhängigkeit der Variablen Y von X unter dieser Konstanzbedingung dieselbe ist, wie diejenige, die man beobachtet, wenn W nicht konstant ist. Von einer intern validen Abhängigkeit der Variablen Y von X erwarten wir also, daß sie sich mit der Ausprägung einer dritten Variablen W nicht verändert, falls W der Variablen X gleich- oder vorgeordnet ist. Eine solche Veränderung war im Beispiel ‚Münzen und Elektromagnet‘ eingetreten, denn die zwischen den Münzvariablen  $Z_I$  und  $Z_{II}$  festgestellte stochastische Abhängigkeit (siehe Tabelle 2.2.3) weicht einer Unabhängigkeit, wenn die Magnetvariable  $Z_M$  konstant ist (siehe die Tabellen 2.2.1 und 2.2.2). Folglich ist die Abhängigkeit der ersten Münzvariablen  $Z_I$  von der zweiten Münzvariablen  $Z_{II}$  nicht intern valide. Da wir interne Validität, neben der zeitlichen Geordnetheit der Variablen, als eine notwendige Bedingung kausaler stochastischer Abhängigkeit ansehen, ist im Fall der Abhängigkeit der ersten von der zweiten Münzvariablen, die durch Gleichung 2.3.4 beschrieben wird, auch keine kausale Interpretation des Steigungskoeffizienten möglich.

Eine ähnliche Auffassung kausaler stochastischer Abhängigkeit scheint auch Suppes (1970, S. 10) zu vertreten, wenn er schreibt: „... ein Ereignis ist die Ursache eines anderen, wenn auf das Erscheinen des ersten Ereignisses mit hoher Wahrscheinlichkeit ein zweites Ereignis folgt, und es kein drittes Ereignis gibt, welches den probabilistischen Zusammenhang zwischen dem ersten und dem zweiten Ereignis zum Verschwinden bringt“ (Übersetzung durch den Autor). Auch bei Suppes sind wohl die beiden wesentlichen Komponenten die zeitliche Abfolge („folgt auf“) und die ideale Situation, daß es nämlich kein drittes Ereignis gibt, welches die Beziehung zwischen den betrachteten beiden Ereignissen „zum Verschwinden bringt“.

In der hier vertretenen Version dieser Grundidee wird diese zweite Bedingung in einer Hinsicht etwas gelockert, indem wir nur von den X gleich- oder vorgeordneten Variablen W verlangen, daß sie den Zusammenhang zwischen X und Y nicht modifizieren und damit verfälschen. In einer zweiten Hinsicht wird diese Bedingung verschärft, indem wir nicht nur keine dritten Ereignisse oder Variablen zulassen, die den Zusammenhang zwischen X und Y „zum Verschwinden“ bringen, sondern jede Art von Modifizierung durch eine solche Variable W verbieten. Die hohe Wahrscheinlichkeit, mit der das zweite Ereignis auf das erste folgen soll, ist dagegen m.E. keine notwendige Bedingung für eine kausale stochastische Abhängigkeit, sondern nur für deren prakti-

tische Bedeutsamkeit. Während in diesem grundlegenden Ansatz große Gemeinsamkeit mit der Arbeit von Suppes festzustellen ist, weichen die hier vorgestellte Theorie und die von Suppes in der Präzisierung dieser Vorstellungen stark voneinander ab, was im einzelnen jedoch hier nicht weiter verfolgt werden kann. Als wichtigste Idee der in beiden Theorien vorkommenden idealen Situation, in der sich eine kausale Abhängigkeit entsprechend äußern kann, und die man als interne Validität bezeichnen kann, bleibt jedoch festzuhalten, daß in einer solchen Situationsklasse die stochastische Abhängigkeit einer Variablen Y von X in gewisser Weise invariant bleibt, wenn eine andere, X gleich- oder vorgeordnete Variable W auf einem bestimmten Wert konstant ist. Dies ist wohl, mit Variablen anstatt mit Ereignissen ausgedrückt, eine ähnliche Vorstellung, wie sie Suppes im obigen Zitat äußert.

Dabei ist allerdings zu beachten, daß wir mit dieser Invarianzbedingung jeweils nur stochastische Variablen auf dem gleichen Wahrscheinlichkeitsraum betrachten, daß wir uns also jeweils nur innerhalb einer bestimmten Untersuchung, einer bestimmten Situationsklasse oder eines bestimmten Experiments bewegen. Selbstverständlich wird beispielsweise die Abhängigkeit der Münzvariablen  $Z_1$  von der Magnetvariablen  $Z_M$  eine ganz andere sein, wenn wir z.B. die Wurfhöhe verändern, so daß die Zeit, in welcher der Elektromagnet wirken kann, verändert ist. Damit betrachten wir jedoch bereits ein anderes Experiment, nämlich eines mit anderer Wurfhöhe, als das in Abschnitt 2 beschriebene, was durch einen anderen Wahrscheinlichkeitsraum zu repräsentieren wäre. Natürlich können wir auch ein größeres Experiment betrachten, in dem beide oder auch beliebig viele Wurfhöhen vorkommen, so daß dann auch der Einfluß des Elektromagneten auf die Münzwurfergebnisse bei verschiedenen Wurfhöhen untersucht werden könnte. Aussagen über die Abhängigkeit der Münz- von der Magnetvariablen in einem solchen größeren Experiment bzw. größeren Wahrscheinlichkeitsraum haben natürlich einen größeren Geltungsbereich - eben diesen größeren Wahrscheinlichkeitsraum - als wenn die Aussage nur für eine einzige Wurfhöhe gemacht wird. Die Frage der Größe des Geltungsbereichs ist von der Frage der internen Validität für einen bestimmten Geltungsbereich verschieden. Erstere hängt eng mit dem Begriff der ‚externen Validität‘ zusammen, auf die wir im Abschnitt 7 näher eingehen. Das Entscheidende beim Konzept der internen Validität ist, daß man sich jeweils nur auf einen bestimmten Geltungsbereich, ein einziges, aber prinzipiell wiederholbares Experiment, eine einzige Situationsklasse bezieht. Diese Aussage behält auch für solche Situationsklassen ihre Gültigkeit, die sehr kompliziert sind, und in denen daher sehr viele Variablen involviert sind.

In den nachfolgenden Abschnitten wollen wir an den aus der Pfadanalyse (siehe z.B. Wright 1921, 1934, 1960a,b) bekannten Pfeilschemata einige Fälle behandeln, in denen eine dritte Variable die interne Validität der Beziehung zwischen zwei Variablen verhindert, und solche, in denen dies nicht zutrifft. Dabei betrachten wir nur solche Fälle, in denen X mit keiner anderen gleich-

oder vorgeordneten Variablen W in einer varianzanalytischen Interaktion steht. Diese besagt ja definitionsgemäß, daß der ‚Effekt‘ von X auf Y nicht auf allen Ausprägungen von W der gleiche ist. Die Variable W würde daher in einem solchen Fall den Zusammenhang zwischen X und Y modifizieren, was gemäß der oben erläuterten Invarianzbedingung gegen eine interne Validität der Abhängigkeit der Variablen Y von X sprechen würde. Im folgenden werden wir uns daher mit einer anderen Art der Modifizierung einer Abhängigkeit zweier Variablen durch eine dritte beschäftigen.

### 3.3 Fälle, in denen keine interne Validität besteht

In den nun zu besprechenden Abbildungen nehmen wir an, daß nur die Variablen X und Y beobachtet werden, daß es aber noch eine dritte Variable W gibt, von deren Existenz der Beobachter aber nichts weiß, was durch die dunkel markierte Fläche angedeutet werden soll. Die durch die Pfeile dargestellten kausalen Abhängigkeiten sind nicht deterministisch, was durch die unbeschrifteten kleinen Pfeile symbolisiert wird.

Abbildung 3.3.1 zeigt eine ähnliche Situation, wie sie bereits im Beispiel ‚Münzen und Elektromagnet‘ beschrieben wurde, in dem nämlich die zwischen den Münzvariablen auftretende stochastische Abhängigkeit durch die Magnetvariable zu erklären ist. In einer solchen Situation besteht keine interne Validität bezüglich der Abhängigkeit der Variablen Y von der Variablen X, da die zwischen X und Y zu beobachtende stochastische Abhängigkeit zumindest teilweise, wenn nicht sogar - wie bei der Abhängigkeit der ersten von der zweiten Münzvariablen - ausschließlich, durch die gemeinsame Abhängigkeit von W zustandekommt. Der Regressionskoeffizient würde in einer solchen Situation nicht das korrekte Ausmaß der einfachen kausalen reglinearen Abhängigkeit angeben.

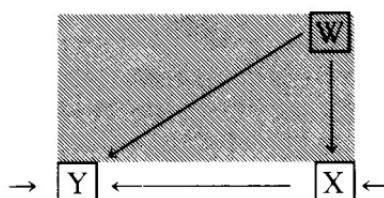


Abb. 3.3.1: In dieser Situation, in der Y von X sowie X und Y von einer dritten Variablen W direkt kausal abhängig sind, besteht bezüglich der Abhängigkeit der Variablen Y von X keine interne Validität.

Simon (1954) hat die in der Abbildung 3.3.1 skizzierte Abhängigkeit der Variablen Y von X als „spurious correlation“ bezeichnet, was mit Scheinkorrelation übersetzt wird (siehe z.B. Hummell & Ziegler, 1976a). Dieser Begriff, der

auch von anderen Autoren (siehe z.B. Namboodiri, Carter & Blalock, 1975, oder Suppes, 1970) verwendet wird, bezeichnet demnach den Fall, daß die betreffende korrelative Abhängigkeit nicht der kausalen Abhängigkeit entspricht, oder daß keine interne Validität gegeben ist.

Wenn wir uns einen Vorgriff auf erst noch einzuführende Begriffe erlauben, so können wir den in Abbildung 3.3.1 dargestellten Fall auch wie folgt formulieren: Wir nehmen an, daß sich die kausalen reglinearen Abhängigkeiten der Variablen Y in der Situation der Abbildung 3.3.1 durch die Gleichung<sup>8)</sup>

$$(3.3.1) \quad E(Y | X, W) \underset{fs}{=} \alpha_{Y0} + \alpha_{YX} X + \alpha_{YW} W$$

beschreiben lassen, wobei  $\alpha_{YX}$  der direkte kausale reglineare Effekt von X auf Y und  $\alpha_{YW}$  der von W auf Y ist. Dabei setzen wir voraus, daß Y den beiden Variablen X und W zeitlich nachgeordnet ist, und daß keine Interaktion im varianzanalytischen Sinn zwischen X und W besteht; andernfalls müßte nämlich die Summe auf der rechten Seite dieser Gleichung noch um den Term  $X \cdot W$ , versehen mit einem entsprechenden Koeffizienten, erweitert werden (siehe z.B. Searle, 1971, S. 289).

Stellen wir in dieser Situation die Frage nach der internen Validität der Abhängigkeit der Variablen Y von X, so wollen wir wissen, ob der Steigungskoeffizient  $\beta_{YX}$  der Gleichung

$$(3.3.2) \quad E(Y | X) \underset{fs}{=} \beta_{Y0} + \beta_{YX} X$$

mit dem Koeffizienten  $\alpha_{YX}$  aus Gleichung 3.3.1 identisch ist, d.h. wir stellen uns damit die Frage, ob es in dieser Situation nötig ist, die Variable W zu kontrollieren, sei es durch ‚statistische Konstanthaltung‘ oder durch ‚faktische Konstanthaltung‘ (siehe Abschnitt 5.3), um den kausalen Effekt  $\alpha_{YX}$  von X auf Y feststellen zu können.

Würden wir nun W ignorieren und die bedingte Erwartung  $E(Y | X)$  bilden, dann erhielten wir aus der Gleichung 3.3.1, unter Verwendung der Gleichungen A.5.7, A.5.10 und A.5.5 des Anhangs

$$(3.3.3) \quad \begin{aligned} E[E(Y | X, W) | X] &\underset{fs}{=} E(Y | X) \underset{fs}{=} \\ &= \alpha_{Y0} + \alpha_{YX} X + \alpha_{YW} E(W | X). \end{aligned}$$

Diese Gleichung würde dann zu einem Steigungskoeffizienten führen, der mit  $\alpha_{YX}$  identisch ist, wenn

$$(3.3.4) \quad \alpha_{YW} = 0,$$

---

<sup>8)</sup> Man beachte die Fußnote zur Gleichung 1.2.1. Bei Referenzen mit drei Ziffern weisen die ersten beiden auf den Abschnitt hin.

oder wenn die bedingte Erwartung

$$(3.3.5) \quad E(W | X)_{\text{fs}} = E(W)$$

konstant wäre. Die letzte Gleichung würde z.B. im Fall der stochastischen Unabhängigkeit von  $X$  und  $W$  zutreffen, und die vorletzte, wenn der direkte kausale reglineare Effekt von  $W$  auf  $Y$  gleich Null wäre. Diese beiden Bedingungen sind aber in dem in Abbildung 3.3.1 dargestellten Fall voraussetzungsgemäß nicht erfüllt. „Interne Validität“ bezüglich der Abhängigkeit der Variablen  $Y$  von  $X$  soll also in dem hier behandelten Fall beinhalten, daß der Term  $\alpha_{YW} E(W | X)$  eine Konstante ist, denn nur dann erhalten wir in Gleichung 3.3.2 einen Steigungskoeffizienten  $\beta_{YX}$ , der mit dem direkten kausalen reglinearen Effekt  $\alpha_{YX}$  von  $X$  auf  $Y$  aus Gleichung 3.3.1 identisch ist, und nur dann würde  $W$  nicht den Zusammenhang zwischen  $X$  und  $Y$  modifizieren oder verfälschen (vgl. Bredenkamp, 1980, S. 1).

Mit diesen Ausführungen wird bereits angedeutet, daß interne Validität bezüglich der Abhängigkeit der Variablen  $Y$  von einer Variablen  $X$  nicht die einzige Bedingung ist, um einen kausalen Effekt von  $X$  auf  $Y$  feststellen zu können. Vielmehr kann man kausale Parameter auch durch „statistische Kontrolle“ erhalten, nämlich durch die Einbeziehung einer weiteren Variablen  $W$  in das Modell (siehe Gleichung 3.3.1). Darauf werden wir etwas ausführlicher im Abschnitt 8 eingehen.

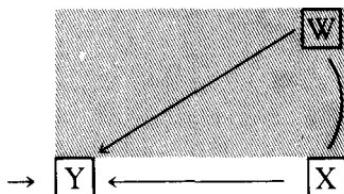


Abb. 3.3.2: In dieser Situation, in der  $Y$  außer von  $X$  auch von einer dritten Variablen  $W$  direkt kausal abhängig ist, die ihrerseits mit  $X$  korreliert, besteht bezüglich der Abhängigkeit der Variablen  $Y$  von  $X$  ebenfalls keine interne Validität.

Eine ganz ähnliche Situation wie die oben besprochene ist auch in Abbildung 3.3.2 dargestellt, in der ebenfalls eine dritte Variable  $W$  Information über  $Y$  enthält, die nicht bereits in  $X$  enthalten ist. Auch für diese Situation nehmen wir an, daß sich die direkten kausalen reglinearen Abhängigkeiten der Variablen  $Y$  durch die Gleichung 3.3.1 beschreiben lassen, und entsprechend sind die gleichen Argumente wie oben auch hier gültig. Im Gegensatz zur Abbildung 3.3.1 ist hier jedoch nur angenommen, daß  $X$  und  $W$  miteinander korrelieren. In einem solchen Fall ist die interne Validität aus den gleichen Gründen wie in der Situation der Abbildung 3.3.1 nicht gegeben, so daß man auch hier nicht davon ausgehen kann, daß die beobachtete reglineare Abhängigkeit der Variablen  $Y$  von  $X$  der kausalen entspricht. Bei Ignorierung von  $W$  ergibt sich

der Regressionskoeffizient  $\beta_{YX}$ , der nicht mit dem kausalen Effekt  $\alpha_{YX}$  aus Gleichung 3.3.1 identisch ist, da wir voraussetzen, daß weder Gleichung 3.3.4 noch Gleichung 3.3.5 (wegen der Korrelation zwischen X und W) gilt. Demnach wird auch in diesem Fall die Beziehung zwischen X und Y durch W modifiziert.

### 3.4 Fälle, in denen möglicherweise interne Validität besteht

In den nun zu besprechenden Fällen spricht die Variable W nicht dagegen, daß die Abhängigkeit der Variablen Y von X intern valide ist, was bei entsprechender zeitlicher Geordnetheit der Variablen heißt, daß W in dieser Situation kein Hindernis ist, die beobachtete reglineare Abhängigkeit der Variablen Y von X kausal zu interpretieren. Dabei beachte man aber, daß man im allgemeinen nicht an einer einzigen potentiellen Störvariablen W entscheiden kann, ob tatsächlich interne Validität vorliegt. Eine einzige Störvariable genügt gegebenenfalls lediglich zur Entscheidung, daß interne Validität nicht vorliegt. Eine positive definitive Entscheidung, daß interne Validität vorliegt, wäre nur möglich, wenn man alle potentiellen Störvariablen kennen und kontrollieren würde. Demnach bleibt in der Regel nur der Weg der Falsifizierung im Sinne Poppers (Popper, 1969).

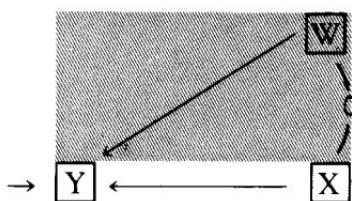


Abb. 3.4.1: In dieser Situation, in der Y außer von X auch von W direkt kausal abhängig ist, spricht W nicht gegen die interne Validität bezüglich der Abhängigkeit der Variablen Y von X. Im Gegensatz zu Abbildung 3.3.2 ist hier angenommen, daß X und W stochastisch unabhängig sind.

In Abbildung 3.4.1 ist dargestellt, daß Y auch von einer dritten Variablen W kausal abhängig ist, wobei jedoch W und X stochastisch unabhängig sind, und außerdem angenommen wird, daß zwischen X und W keine Interaktion im varianzanalytischen Sinn besteht, so daß die Abhängigkeit der Variablen Y von X bei verschiedenen Ausprägungen von W identisch ist. Wegen der hier vorausgesetzten Unabhängigkeit zwischen X und W aber läßt sich die Abhängigkeit zwischen Y und X nicht durch W erklären, da die stochastische Unabhängigkeit zwischen X und W die Gleichung 3.3.5 impliziert, so daß der Term  $\alpha_{YW} E(W | X)$  in Gleichung 3.3.3 eine Konstante ist. Daraus folgt aber, daß die Koeffizienten  $\alpha_{YX}$  und  $\beta_{YX}$  aus den Gleichungen 3.3.1 bzw. 3.3.2 identisch

sind. In dieser Situation modifiziert also W die Beziehung zwischen X und Y nicht, d.h. W ist keine konfundierende Variable.

Im Fall der Abbildung 3.4.2 ist X selbst von einer weiteren Variablen W direkt kausal abhängig, aber Y ist von W direkt kausal reglinear unabhängig, was durch den Nullpfeil von W nach Y symbolisiert ist. Auf Gleichung 3.3.1 bezogen heißt das, daß  $\alpha_{YW} = 0$  gilt, weswegen auch in diesem Fall der Term  $\alpha_{YW} E(W | X)$  in Gleichung 3.3.3 eine Konstante ist, so daß auch hier W nicht den Zusammenhang zwischen X und Y modifiziert. Dabei beachte man allerdings, daß wir voraussetzen, daß zwischen X und W keine Interaktion im varianzanalytischen Sinn besteht.

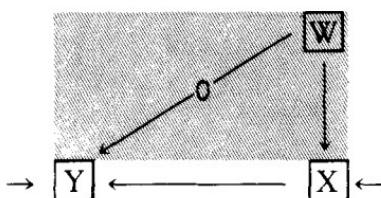


Abb. 3.4.2: In dieser Situation ist X selbst von einer dritten Variablen W direkt kausal abhängig, so daß auf diesem Weg Y und W zwar stochastisch voneinander abhängig sind, aber nicht direkt kausal. In dieser Situation spricht W nicht gegen die interne Validität bezüglich der Abhängigkeit der Variablen Y von X.

Im Fall der Abbildung 3.4.3 ist die kausale Abhängigkeit zwischen Y und X lediglich nicht mehr direkt, wenn man von der um W erweiterten Variablenmenge ausgeht. In diesem Punkt unterscheidet sich die hier entwickelte Auffassung interner Validität von der von Cook und Campbell (1979, S. 50)

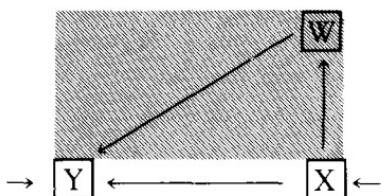


Abb. 3.4.3 In dieser Situation ist sowohl Y als auch W von der Variablen X direkt kausal abhängig. Darüber hinaus ist Y auch von W direkt kausal abhängig. Die interne Validität der Abhängigkeit der Variablen Y von X ist durch W nicht gefährdet. Die bezüglich der Variablenmenge  $\{X,Y\}$  direkte kausale Abhängigkeit der Variablen Y von X ist jedoch bezogen auf die um W erweiterte Menge  $\{W,X,Y\}$  nur noch eine totale, die sich aus direkten und indirekten kausalen Abhängigkeiten zusammensetzt.

vertretenen, da die genannten Autoren in diesem Fall, anstelle der Direktheit der Beziehung, deren interne Validität in Frage stellen. In der Regel können wir jedoch nie ausschließen, daß es, wie in Abbildung 3.4.3 skizziert, noch weitere Variablen W gibt, die zwischen X und Y anzutreffen wären. Im Falle der zeitlichen Interpretation einer solchen Ordnung heißt dies, daß es wohl immer eine zwischen der beeinflussenden und der beeinflußten Variablen anzutreffende dritte Variable W gibt, auch wenn diese bis dahin unbekannt und nicht erhoben ist. Ohne die Existenz solcher zwischen Ursache und Wirkung vermittelnden Variablen wäre ein kausaler Zusammenhang wohl auch kaum denkbar. Dabei könnte sogar der Fall eintreten, daß überhaupt keine direkte kausale Abhängigkeit der Variablen Y von X besteht und deren beobachtete stochastische Abhängigkeit allein über eine vermittelnde Variable W zu erklären ist.

Als Beispiel für die in Abbildung 3.4.3 dargestellte Situation kann Rosenthal's (siehe z.B. Rosenthal, 1964, 1966, Selg & Bauer, 1976, S. 59ff. oder auch Timäus, 1974) Versuchsleitereffekt angeführt werden, bei dem die oft unberücksichtigte Variable W = „Erwartung des Versuchsleiters“ einen Einfluß auf die abhängige Variable Y hat und ihrerseits von der eigentlich unabhängigen Variablen X beeinflußt wird, welche die Zugehörigkeit zu den experimentellen Gruppen anzeigt.

Bei einem Versuchsleitereffekt handelt es sich also um einen Unterschied zwischen den experimentellen Gruppen, oder, was damit gleichbedeutend ist, um einen Zusammenhang zwischen der abhängigen Variablen und der unabhängigen Variablen, welche die Gruppenzugehörigkeit anzeigt, der nicht direkt auf die eigentlichen experimentellen Behandlungen zurückzuführen ist, sondern auf die vom Versuchsleiter „ausgestrahlten“ Erwartungen bezüglich des Versuchsergebnisses, welche die Versuchspersonen bewußt oder unbewußt zu erfüllen bestrebt sind. Solchen Effekten ist nur durch Blind- oder Doppelblindversuche vorzubeugen, bei denen auch der Versuchsleiter im unklaren über den Zweck des Versuchs gelassen wird.

Ein weiteres Beispiel für die Situation der Abbildung 3.4.3 ist der sogenannte Hawthorneeffekt, den Wilkening und Wilkening (1979, S. 15) wie folgt beschreiben: „In einem der Experimente sollte die Leistung von Arbeitern an einem besonders hell beleuchteten Arbeitsplatz („Experimentalgruppe“) mit der Leistung von Arbeitern an einem normal beleuchteten Platz („Kontrollgruppe“) verglichen werden. Die Experimentalgruppe zeigte höhere Leistungswerte. Aufgrund dieses Ergebnisses schloß der Versuchsleiter, daß zusätzliche Beleuchtung die Produktivität erhöht.“

In einem Anschlußexperiment wurde die Beleuchtung bei der Experimentalgruppe wieder derjenigen der Kontrollgruppe angeglichen. Erstaunlicherweise verschlechterte sich hierdurch die Arbeitsleistung der Experimentalgruppe nicht.“

Diese Ergebnisse interpretieren Wilkening und Wilkening folgendermaßen: „Aufgrund der Versuchsbeschreibung ist anzunehmen, daß eine wesentliche Variable nicht kon-

stantgehalten worden ist, und zwar das Wissen der Arbeiter, Teilnehmer an einem Experiment zu sein. Während die Mitglieder der „Kontrollgruppe“ nichts von einer im Werk durchgeföhrten Untersuchung wußten, fühlten sich die Arbeiter der Experimentalgruppe dadurch, daß ihr Verhalten beobachtet und registriert wurde, als „Versuchspersonen“, die sich der Aufmerksamkeit der Geschäftsleitung bewußt waren. Dieses Wissen allein genügte, um die Leistung zu verbessern. Hätte man es konstantgehalten, indem man die Mitglieder von Experimental- und Kontrollgruppe in gleicher Weise über ihre Rolle als Versuchsperson informiert hätte, wären vermutlich keine Unterschiede zwischen den beiden Beleuchtungsbedingungen beobachtet worden“ (1979, S.t7).

Offenbar wurde bei dieser Untersuchung die Variable  $W = \text{„Motivation durch Wissen/Nichtwissen um Teilnahme an einer Untersuchung“}$  von der eigentlichen unabhängigen Variablen  $X$  beeinflußt, so daß die festgestellte Abhängigkeit der Variablen  $Y$  von  $X$  nicht direkt auf  $X$  zurückgeführt werden kann, wenn man dieses „direkt“ auf die Variablenmenge  $\{W, X, Y\}$  bezieht. Bezüglich der beiden Variablen  $X$  und  $Y$  wäre diese Abhängigkeit hingegen direkt.

Die in Abbildung 3.4.3 dargestellten kausalen Abhängigkeiten lassen sich durch die beiden Gleichungen

$$(3.4.1) \quad E(Y | X, W) \underset{\text{fs}}{=} \alpha_{Y0} + \alpha_{YX} X + \alpha_{YW} W, \quad \text{mit } \alpha_{YW} \neq 0,$$

und

$$(3.4.2) \quad E(W | X) \underset{\text{fs}}{=} \alpha_{W0} + \alpha_{WX} X, \quad \text{mit } \alpha_{WX} \neq 0,$$

beschreiben. Würden wir in diesem Fall  $W$  ignorieren, so erhielten wir als Steigungskoeffizienten nicht  $\alpha_{YX}$ , der die bezüglich  $\{W, X, Y\}$  direkte kausale reglineare Abhängigkeit der Variablen  $Y$  von  $X$  angibt, sondern den totalen kausalen reglinearen Effekt  $\beta_{YX} := \alpha_{YX} + \alpha_{YW} \cdot \alpha_{WX}$  von  $X$  auf  $Y$ , denn nach den Gleichungen A.5.7, A.5.10 und A.5.5 des Anhangs gilt

$$\begin{aligned} (3.4.3) \quad E[E(Y | X, W) | X] &\underset{\text{fs}}{=} E(Y | X) \underset{\text{fs}}{=} \\ &= \alpha_{Y0} + \alpha_{YX} X + \alpha_{YW} E(W | X) \underset{\text{fs}}{=} \\ &= \underset{\text{fs}}{(\alpha_{Y0} + \alpha_{YW} \cdot \alpha_{W0}) + (\alpha_{YX} + \alpha_{YW} \cdot \alpha_{WX}) X}, \end{aligned}$$

wobei man den Übergang von der zweiten zur dritten Formelzeile durch Einsetzen der Gleichung 3.4.2 erhält. Man beachte, daß der bezüglich aller drei Variablen totale kausale reglineare Parameter  $\beta_{YX} := \alpha_{YX} + \alpha_{YW} \cdot \alpha_{WX}$  bezüglich der Variablenmenge  $\{X, Y\}$  direkt ist.

In Abbildung 3.4.4 schließlich ist von  $Y$  noch eine weitere Variable  $W$  abhängig, was für die interne Validität bezüglich der Variablen  $Y$  von  $X$  nicht von Bedeutung ist, denn es gibt wohl keinen Grund zu erwarten, daß bei Konstanz einer nachgeordneten Variablen  $W$  noch die gleiche stochastische Abhängig-

keit zwischen Y und X besteht. Wenn nämlich Y und W stochastisch abhängig sind, so haben wir mit dem Wert von W auch Information über Y. Bei einer vollständigen Abhängigkeit der Variablen W von Y läge sogar mit einem Wert von W auch der von Y fest, so daß dann bei gegebenem Wert von W die Variable Y konstant und daher unabhängig von X wäre. Dies aber sollte wohl nicht gegen eine eventuelle interne Validität der Abhängigkeit der Variablen Y von X sprechen.

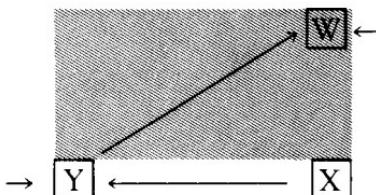


Abb. 3.4.4: In dieser Situation ist nicht nur Y von X, sondern auch die dritte Variable W von Y direkt kausal abhängig. Dadurch wird die interne Validität der Abhängigkeit der Variablen Y von X nicht gefährdet.

Bei allen in diesem Abschnitt behandelten Fällen haben wir vorausgesetzt, daß die Abbildungen die tatsächlichen kausalen Abhängigkeiten beschreiben, daß also die in diesen Pfeildiagrammen angedeuteten Fehlervariablen „strikt zufällig“ (vgl. Namooddiri et al., 1975, S. 446) sind. Auf diese Weise haben wir zwar einige Grundideen erläutern können, aber das eigentliche Problem der Definition einer ‚kausalen‘ oder ‚intern validen‘ Abhängigkeit ist damit noch nicht gelöst, genausowenig wie das Problem, wie man in Anwendungsfällen empirisch überprüfen kann, ob ein bestimmtes Pfeildiagramm und die zugehörigen Gleichungen tatsächlich die kausalen Abhängigkeiten beschreiben. Diese Aufgaben wollen wir uns in den folgenden Abschnitten zuwenden.

### 3.5 Zusammenfassende Bemerkungen

In diesem Abschnitt wurde die Grundidee interner Validität bezüglich der Abhängigkeit einer Variablen Y von einer zweiten Variablen X erläutert. Diese besteht darin, daß keine dritte X gleich- oder vorgeordnete Variable W existiert, welche die Beziehung zwischen X und Y modifiziert. Gibt es eine dritte Variable W, die zwischen X und Y anzusiedeln ist, dann kann W nicht die interne Validität, sondern nur die Direktheit der Beziehung in Frage stellen. Hierin unterscheidet sich die hier vertretene Auffassung von der von anderen Autoren geäußerten (siehe z.B. Cook & Campbell, 1979, S. 50). Die oben genannte Grundidee wurde durch die folgende Invarianzbedingung konkretisiert: Die Beziehung zwischen X und Y muß bei Konstanz einer X gleich- oder

vorgeordneten Variablen W invariant bleiben, wenn interne Validität bezüglich der Abhängigkeit der Variablen Y und X bestehen soll. Anhand verschiedener Pfeilschemata wurde diese Vorstellung erläutert.

## 4. Einfache kausale reglineare Abhängigkeit

### 4.1 Einleitende Bemerkungen

In diesem Abschnitt ist es unser Ziel, den Begriff ‚einfache kausale reglineare Abhängigkeit‘ formal zu definieren, und damit die idealen Bedingungen anzugeben, unter denen eine kausale regressiv lineare oder ‚reglineare‘ Abhängigkeit zwischen einer Variablen Y und *einer* Variablen X beobachtet werden kann, um im nächsten Abschnitt die Eigenschaften dieses Begriffs mathematisch untersuchen zu können. Diese sollen dann angeben, in welchen Fällen eine einfache kausale reglineare Abhängigkeit vorliegt, und in welchen nicht.

Wir sprechen dabei von einfacher kausaler reglinearer Abhängigkeit, um diese Art der kausalen Abhängigkeit, die sich durch eine *einfache* Regressionsgleichung mit einer unabhängigen Variablen beschreiben lässt, von der direkten kausalen reglinearen Abhängigkeit abzugrenzen, zu deren Beschreibung eine multiple Regressionsgleichung mit mehreren unabhängigen Variablen benötigt wird. Der Begriff der direkten kausalen reglinearen Abhängigkeit wird dann wichtig, wenn die Situation eine (statistische) Kontrolle anderer unabhängiger Variablen erfordert, die möglicherweise auch zwischen X und Y angeordnet sind.

In den folgenden Abschnitten besprechen wir zunächst die Bedingungen der Vorgeordnetheit sowie der Invarianz getrennt, und fassen diese dann in der Definition einfacher kausaler reglinearer Abhängigkeit zusammen.

### 4.2 Vorgeordnetheit

Eine Formalisierung der im Abschnitt 3 beschriebenen intuitiven Vorstellungen interner Validität ist offenbar nur möglich, wenn wir in einer gegebenen Situationsklasse von den Variablen X und Y entscheiden können, ob X der Variablen Y vorgeordnet ist. In Anwendungen kann man sich als dabei zu verwendendes Kriterium ‚X ist zeitlich Y vorgeordnet‘ denken. Eine gewisse Geordnetheit der Variablen ist aber auch aus folgenden Gründen wichtig: Existiert eine weitere Variable W, die X gleich- oder vorgeordnet ist (siehe die Abbildungen 3.3.1 bis 3.4.2), dann ist W eine Variable, welche möglicherweise die interne Validität der Abhängigkeit der Variablen Y von X gefährden kann, wie z.B. in den Abbildungen 3.3.1 und 3.3.2. Ist hingegen X einer dritten

Variablen W vorgeordnet, wie in den Abbildungen 3.4.3 und 3.4.4, dann ist die interne Validität der Abhängigkeit der Variablen Y von X durch W nicht gefährdet, sondern nur deren Direktheit.

Wir wollen nun präzisieren, was wir unter Vorgeordnetheit verstehen wollen. Zur Veranschaulichung der dazu benötigten Begriffe der Stochastik betrachten wir zunächst den Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$ , der die bestimmte Situationsklasse repräsentieren soll, in der die Münzen I und II geworfen werden. Dabei besteht die Menge

$$(4.2.1) \quad \Omega = \{(a_1, b_1), (a_1, b_2), (a_2, b_1), (a_2, b_2)\},$$

aus den vier Elementarereignissen

$$(4.2.2) \quad (a_i, b_j) = \begin{array}{l} \text{Münze I fällt auf die i-te Seite und} \\ \text{Münze II fällt auf die j-te Seite,} \end{array} \quad i, j \in \{1, 2\},$$

die beim Werfen der beiden Münzen I und II auftreten können. Nun können wir die mit dem Werfen der ersten Münze verbundenen Ereignisse in der Sigmaalgebra

$$(4.2.3) \quad \{\{(a_1, b_1), (a_1, b_2)\}, \{(a_2, b_1), (a_2, b_2)\}, \Omega, \emptyset\}$$

zusammenfassen, die zugleich die von der stochastischen Variablen

$$(4.2.4) \quad Z_I = \begin{cases} 1, & \text{wenn die Münze I auf die Seite 1,} \\ 0, & \text{wenn sie auf die Seite 2 fällt,} \end{cases}$$

erzeugte Sigmaalgebra  $\mathcal{A}(Z_I)$  ist (vgl. z.B. Bauer, 1974, S. 42). Dabei repräsentiert beispielsweise  $A = \{(a_1, b_1), (a_1, b_2)\}$  das Ereignis, daß die Münze I auf die Seite 1 fällt und daß  $Z_I$  den Wert 1 annimmt.

Die von einer Funktion und im besonderen von einer stochastischen Variablen X erzeugte Sigmaalgebra enthält alle Ereignisse

$$(4.2.5) \quad A := \{\omega \in \Omega : X(\omega) = x\},$$

daß X einen bestimmten Wert x angenommen hat, als Elemente, aber sie enthält auch alle Vereinigungsmengen solcher Ereignisse. Die von der stochastischen Variablen

$$(4.2.6) \quad Z_{II} = \begin{cases} 1, & \text{wenn die Münze II auf die Seite 1,} \\ 0, & \text{wenn sie auf die Seite 2 fällt,} \end{cases}$$

erzeugte Sigmaalgebra  $\mathcal{A}(Z_{II})$  ist daher

$$(4.2.7) \quad \{\{(a_1, b_1), (a_2, b_1)\}, \{(a_1, b_2), (a_2, b_2)\}, \Omega, \emptyset\}.$$

Diese Formalisierung ist jedoch für unsere Zwecke zu vereinfachend, da die beim Werfen zweier Münzen auftretenden Ereignisse nicht nur die unter 4.2.3 und 4.2.7 genannten sind, die ja nur die Endergebnisse des Werfens der beiden Münzen repräsentieren. Auch bei einem solch einfachen Beispiel könnten wir zu jedem Zeitpunkt  $t$  in diesem Versuch unendlich viele Ereignisse nennen, z.B. jene, welche die räumliche Lage der Münzen oder deren Drehgeschwindigkeit angeben. Eine vollständige explizite Beschreibung dieses Versuchs müßte alle diese Ereignisse enthalten, einschließlich der diesen Ereignissen zugeordneten Wahrscheinlichkeiten. Das bedeutet jedoch nicht, daß wir alle Ereignisse explizit angeben oder gar, daß deren Wahrscheinlichkeiten bekannt sein oder geschätzt werden müssen. Allerdings muß eine formale Struktur zugrunde gelegt werden, die uns zum einen in die Lage versetzt, bei der formalen Abhandlung zwischen je zwei Zeitpunkten  $s$  und  $t$  einen weiteren Zeitpunkt anzunehmen, und zum anderen ermöglicht, gegebenenfalls auf alle bei diesem Versuch möglichen Ereignisse und Variablen zurückgreifen zu können, die ja u.U. Störereignisse bzw. Störvariablen sein können.

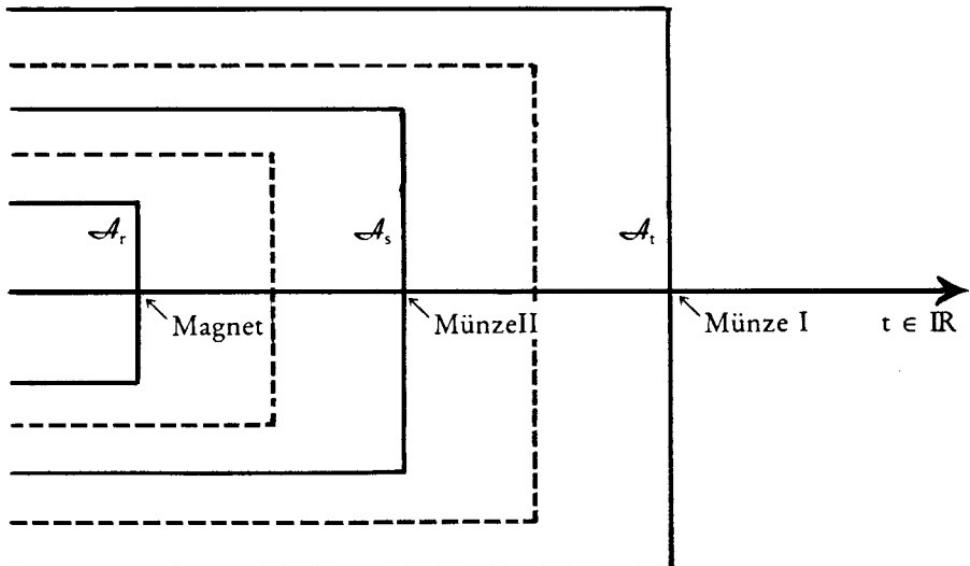


Abb. 4.2.1: Anordnung der betrachteten Ereignisse auf der Zeitachse und schematisierte Darstellung einer isotonen Familie ( $\mathcal{A}_t, t \in \mathbb{IR}$ ) von Sigmaalgebren.

Eine formale Struktur, die diesen Anforderungen genügt, ist die isotone oder monoton wachsende Familie von Sigmaalgebren, worunter man eine Familie  $(\mathcal{A}_t, t \in \mathbb{IR})$  von Sigmaalgebren versteht mit der Eigenschaft

$$(4.2.8) \quad \mathcal{A}_s \subset \mathcal{A}_t, \text{ falls } s \leqq t, \text{ wobei } s, t \in \mathbb{IR}$$

(siehe z.B. Bauer, 1974, S. 314). Demnach enthält eine solche Sigmaalgebra

$\mathcal{A}_t$  alle bis zum Punkt  $t$  möglichen Ereignisse als Elemente (siehe Abbildung 4.2.1).

Indem wir eine isotone Familie von Sigmaalgebren zugrunde legen, können wir ‚Vorgeordnetheit‘ sowohl für stochastische Variablen als auch für Sigmaalgebren und Ereignisse nun wie folgt definieren:

*Definition 4.2.1.* Es seien  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum,  $\mathcal{B}$  und  $\mathcal{C}$  zwei Teilsigmaalgebren von  $\mathcal{A}$ ,  $\mathbb{R}$  die Menge der reellen Zahlen versehen mit den üblichen Ordnungsrelationen  $\geq, \leq, <$  sowie  $>$ , und  $(\mathcal{A}_t, t \in \mathbb{R})$  sei eine isotone Familie von Sigmaalgebren mit

$$(4.2.9) \quad \mathcal{A}(\bigcup_{t \in \mathbb{R}} \mathcal{A}_t) = \mathcal{A}.$$

Die Sigmaalgebra  $\mathcal{B}$  heißt  $\mathcal{C}$  (bezüglich  $(\mathcal{A}_t, t \in \mathbb{R})$ ) vorgeordnet genau dann, wenn zwei Elemente  $s, t \in \mathbb{R}$ ,  $s < t$ , existieren, für die gilt:

$$(4.2.10) \quad \mathcal{B} \subset \mathcal{A}_s, \mathcal{C} \subset \mathcal{A}_t \text{ und } \mathcal{C} \not\subset \mathcal{A}_s.$$

Die stochastische Variable  $X$  heißt der stochastischen Variablen  $Y$  (bezüglich  $(\mathcal{A}_t, t \in \mathbb{R})$ ) vorgeordnet genau dann, wenn die von  $X$  erzeugte Sigmaalgebra  $\mathcal{A}(X)$  der von  $Y$  erzeugten Sigmaalgebra  $\mathcal{A}(Y)$  vorgeordnet ist.

Das Ereignis  $A \in \mathcal{A}$  heißt dem Ereignis  $B \in \mathcal{A}$  (bezüglich  $(\mathcal{A}_t, t \in \mathbb{R})$ ) vorgeordnet genau dann, wenn die Indikatorvariable  $1_A$  der Indikatorvariablen  $1_B$  vorgeordnet ist<sup>9</sup>.

Auf unser obiges Beispiel des Werfens zweier Münzen angewandt, können wir also sagen, daß die Variable  $Z_{II}$  der Variablen  $Z_I$  vorgeordnet ist, da die Münze II zum Zeitpunkt  $s$  und die Münze I zum Zeitpunkt  $t$  geworfen wird. Enthält die Sigmaalgebra  $\mathcal{A}_s$  alle bis zum Zeitpunkt  $s$  möglichen Ereignisse, so ist also  $\mathcal{A}(Z_{II})$  Teilmenge von  $\mathcal{A}_s$ , nicht jedoch  $\mathcal{A}(Z_I)$ , da Münze I erst zum Zeitpunkt  $t > s$  geworfen wird:

$$(4.2.11) \quad \mathcal{A}(Z_{II}) \subset \mathcal{A}_s, \mathcal{A}(Z_I) \subset \mathcal{A}_t \text{ und } \mathcal{A}(Z_I) \not\subset \mathcal{A}_s$$

(siehe Abbildung 4.2.1).

Auf die entsprechende Weise ist auch die Vorgeordnetheit einer Sigmaalgebra gegenüber einer stochastischen Variablen oder die Vorgeordnetheit einer stochastischen Variablen gegenüber einem Ereignis und umgekehrt definiert. Eine stochastische Variable  $X$  z.B. soll einem Ereignis  $A$  vorgeordnet heißen genau

<sup>9</sup>) Eine stochastische Variable, die nur die Werte Null (für  $\bar{A}$ ) und Eins (für  $A$ ) annehmen kann, nennt man in der Wahrscheinlichkeitstheorie Indikatorvariable für das Ereignis  $A$  und man benutzt normalerweise das Symbol  $1_A$ .

dann, wenn  $\mathcal{A}(X)$  der von den Indikatorvariablen  $1_A$  erzeugten Sigmaalgebra  $\mathcal{A}(1_A) = \{A, \bar{A}, \Omega, \emptyset\}$  vorgeordnet ist.

Der Begriff der Vorgeordnetheit bezieht sich also immer auf eine bestimmte isotone Familie  $(\mathcal{A}_t, t \in \mathbb{R})$  von Sigmaalgebren. In Anwendungen<sup>10</sup> kann eine solche Sigmaalgebra  $\mathcal{A}_t$  als die Menge der bis einschließlich zum Zeitpunkt  $t$  möglichen Ereignisse interpretiert werden. Dabei sind nicht etwa Meßzeitpunkte entscheidend, sondern vielmehr die ‚Wirkzeitpunkte‘. Wann Variablen tatsächlich erhoben werden, ist eher eine technische Frage und hat mit der eigentlich inhaltlichen Theorie nur bedingt zu tun.

Auch der Begriff ‚Gleichgeordnetheit‘ läßt sich nun exakt wie folgt bestimmen :

---

**Definition 4.2.2.** Es mögen wieder die Voraussetzungen und Schreibweisen gelten, die in der Definition 4.2.1 aufgeführt sind.

Die *Sigmaalgebra*  $\mathcal{B}$  heißt  $\mathcal{C}$  (bezüglich  $(\mathcal{A}_t, t \in \mathbb{R})$ ) gleichgeordnet genau dann, wenn  $\mathcal{C}$  nicht  $\mathcal{B}$  und  $\mathcal{B}$  nicht  $\mathcal{C}$  vorgeordnet ist, und ein  $t \in \mathbb{R}$  existiert mit

$$(4.2.12) \quad \mathcal{B} \subset \mathcal{A}_t \text{ und } \mathcal{C} \subset \mathcal{A}_t.$$

Die stochastische Variable  $X$  heißt der stochastischen Variablen  $Y$  (bezüglich  $(\mathcal{A}_t, t \in \mathbb{R})$ ) gleichgeordnet genau dann, wenn  $\mathcal{A}(X)$  und  $\mathcal{A}(Y)$  gleich geordnet sind.

Das Ereignis  $A \in \mathcal{A}$  heißt dem Ereignis  $B \in \mathcal{A}$  (bezüglich  $(\mathcal{A}_t, t \in \mathbb{R})$ ) gleichgeordnet genau dann, wenn  $\mathcal{A}(1_A)$  und  $\mathcal{A}(1_B)$  gleichgeordnet sind.

---

Gleichgeordnetheit kann in Anwendungsfällen als zeitliche Gleichgeordnetheit interpretiert werden, aber andere Interpretationen sind ebensogut denkbar. Als Beispiel zweier gleichgeordneter Variablen kann man sich zwei gleichzeitig experimentell manipulierte unabhängige Variablen  $X_1$  und  $X_2$  vorstellen.

### 4.3 Invarianz

Die Bedingung, daß  $X$  der Variablen  $Y$  vorgeordnet ist, ist eine notwendige, aber keineswegs hinreichende Bedingung einer einfachen kausalen reglinearen

---

<sup>10</sup>) Wie bei jedem formalen Begriff sind auch bei dem der Vorgeordnetheit prinzipiell andere Anwendungen und „Interpretationen“ erlaubt, insbesondere auch solche, in denen  $\mathbb{R}$  nicht als Zeitmenge interpretiert wird. Ob solche alternativen Interpretationen ebenfalls zu etwas Sinnvollem führen, vermag ich zur Zeit nicht zu beurteilen.

Abhängigkeit der Variablen Y von X, wie wir bereits im Abschnitt 2.3 festgestellt haben, in dem die Abhängigkeit der ersten von der zweiten Münzvariablen für das Beispiel ‚Münzen und Elektromagnet‘ behandelt wurde. Unser Ziel ist nun, formale Eigenschaften zu finden, welche die Fälle der Abbildungen 3.3.1 und 3.3.2, in denen keine interne Validität besteht, von denjenigen der Abbildungen 3.4.1 und 3.4.2 unterscheiden, in denen durch die dritte Variable W die Validität nicht in Frage gestellt wird. Für die anderen beiden Fälle der Abbildungen 3.4.3 und 3.4.4 haben wir ja bereits festgestellt, daß dort X der Variablen W vorgeordnet ist, womit W die Validität der Abhängigkeit der Variablen Y von X nicht gefährden kann. In diesen Fällen ist W also keine potentielle Störvariable, da sie höchstens, wie im Fall der Abbildung 3.4.3, dazu führt, daß eine bezüglich  $\{X, Y\}$  direkte kausale reglineare Abhängigkeit, bezogen auf die um W erweiterte Variablenmenge  $\{W, X, Y\}$  nicht mehr direkt ist (vgl. hierzu auch Hummell & Ziegler, 1976b, S. E 57). Zur Unterscheidung der Fälle der Abbildungen 3.3.1 und 3.3.2 von denen der Abbildungen 3.4.1 und 3.4.2 ist offenbar die Forderung, daß X der Variablen Y vorgeordnet ist, nicht hinreichend, da dies für jeden der genannten vier Fälle zutrifft, wenn auch diese Vorgeordnetheit zweifellos eine notwendige Bedingung für eine einfache kausale reglineare Abhängigkeit ist.

Notwendig, aber ebenfalls nicht hinreichend ist die Gültigkeit einer Gleichung vom Typ

$$(4.3.1) \quad E(Y | X)_{fs} = \alpha_{Y0} + \alpha_{YX} X,$$

durch die eine reglineare Abhängigkeit definiert ist. Die in dieser Gleichung vorkommende reelle Zahl bezeichnen wir als Steigungskoeffizient der Regression von Y unter X. In Abschnitt 2 haben wir gesehen, daß mit einer solchen Gleichung sowohl die intuitiv kausale reglineare Abhängigkeit der Münz- von der Magnetvariablen, als auch die nichtkausale reglineare Abhängigkeit der ersten von der zweiten Münzvariablen beschrieben werden können (siehe die Gleichungen 2.3.4, 2.4.2 und 2.4.3).

Diese gesuchte Eigenschaft<sup>11)</sup>, die neben der Vorgeordnetheit eine einfache kausale reglineare Abhängigkeit der Variablen Y von X definieren soll, ist die Formalisierung der Vorstellung, daß eine solche Abhängigkeit nicht von der Ausprägung w einer potentiellen Störvariablen W abhängen darf. Damit präzisieren wir die Idee, daß eine potentielle Störvariable W nicht den Zusammenhang zwischen X und Y „modifizieren“ (Bredenkamp, 1980, S. 1) darf, oder

<sup>11)</sup> Diese Eigenschaft wurde ähnlich auch schon von Lazarsfeld (1955, S. 125) verbal formuliert. Hummell und Ziegler (1976b, S. E 32) haben nicht akzeptiert, daß dabei tatsächlich alle potentiellen Störvariablen auf dem zugrunde gelegten Wahrscheinlichkeitsraum wichtig sind, und eben nicht nur die in der Theorie spezifizierten, die also in der Modellgleichung vorkommen. Das Argument, daß solche Kausalaussagen nicht im Sinne einer Verifikation überprüft werden können, spricht nicht gegen ein solches Konzept, da ja die Überprüfung nach dem Falsifikationsverfahren möglich ist.

anders formuliert, daß keine „konfundierenden Faktoren“ existieren dürfen (vgl. z.B. Sarris, 1968, S. 183).

Dabei ist es unerheblich, ob eine solche Variable W erhoben wurde oder nicht. Wesentlich ist nur, daß W eine stochastische Variable auf dem gleichen Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  ist, der die betreffende Situationsklasse repräsentiert, so daß W prinzipiell erhoben werden könnte. Beim Beispiel ‚Münzen und Elektromagnet‘ beispielsweise ist die Magnetvariable  $Z_M$  auch dann vorhanden und hat eine Wirkung auf die Münzwurfergebnisse, wenn sie nicht erhoben wird. Dagegen wäre eine Variable, die den Zustand eines zweiten Elektromagneten anzeigen würde und eine Varianz größer Null hätte, keine stochastische Variable auf dem gleichen Wahrscheinlichkeitsraum, da aus der Beschreibung des Beispiels zumindest implizit hervorgeht, daß dort nur ein einziger Elektromagnet installiert ist. Ein Experiment, in dem zwei Elektromagneten vorkommen, würde durch einen anderen Wahrscheinlichkeitsraum repräsentiert, als das beschriebene Experiment, in dem es nur einen einzigen Elektromagneten gibt.

Wir wollen nun zunächst präzisieren, was wir unter potentiellen Störvariablen W verstehen wollen. Die Bedingung, daß W der Variablen X gleich- oder vorgeordnet ist, reicht nicht aus, da dann auch  $W_1 = X \cdot Z$ ,  $W_2 = X + Z$  u.ä. potentielle Störvariablen wären, falls Z der Variablen X gleich- oder vorgeordnet ist, was aus Definition 4.2.2 folgt. Als potentielle Störvariablen sollen aber nur solche Variablen gelten, die unabhängig von X definiert sind. Die folgende Definition stellt dies sicher.

**Definition 4.3.1.** Es mögen wieder die Voraussetzungen und Schreibweisen gelten, die in der Definition 4.2.1 aufgeführt sind. Außerdem seien X und W stochastische Variablen und  $Z: \Omega \rightarrow \mathbb{IR}^n$ ,  $n \in \{1, 2, \dots\}$ , ein stochastischer Vektor auf  $(\Omega, \mathcal{A}, P)$ .

W heißt potentielle Störvariable bezüglich X [und  $(\mathcal{A}_t, t \in \mathbb{IR})$ ] genau dann, wenn gilt: a) W ist X gleich- oder vorgeordnet, b) Wenn  $W = f(X, Z)$ , wobei Z der Variablen X gleich- oder vorgeordnet ist und  $f: X(\Omega) \times Z(\Omega) \rightarrow \mathbb{IR}$ , so existiert eine Funktion  $g: Z(\Omega) \rightarrow \mathbb{IR}$  mit  $f(X, Z) = g(Z)$ . [ $X(\Omega)$  ist dabei das Bild von  $\Omega$  unter X.]

Variablen wie  $W_1 = X \cdot Z$  oder  $W_2 = X + Z$  sind demnach keine potentiellen Störvariablen bezüglich X, da sie nicht nur von Z, sondern auch von X abhängen (siehe Punkt b der obigen Definition).

Am folgenden Fall soll nun die Invarianzbedingung weiter erläutert werden: Angenommen, die Variable Y wäre sowohl von X, als auch von W kausal abhängig, und diese kausale Abhängigkeit ließe sich durch

$$(4.3.2) \quad E(Y | X, W) \stackrel{fs}{=} \beta_{Y0} + \beta_{YX} X + \beta_{YW} W$$

beschreiben, wobei X und W stochastisch unabhängig seien, so daß auch  $E(W|X)$

$= E(W)$  gilt. Dann gälte nach den Gleichungen A.5.7, A.5.5 und A.5.10 die Gleichung 4.3.1 mit  $\alpha = \beta_{Y0} + \beta_{YW} E(W)$  und  $\alpha = \beta_{YX}$ , d.h. der Effekt von W auf Y wäre additiv zu dem von X. Für die bedingte Erwartung von Y unter X gegeben das Ereignis A, daß W den Wert w angenommen hat, gilt dann auch

$$(4.3.3) \quad E_A(Y | X) \underset{P_A-f_s}{=} (\beta_{Y0} + \beta_{YW} w) + \beta_{YX} X = \alpha_{Y0} + \alpha_{YX} X,$$

wobei  $\alpha = \beta_{Y0} + \beta_{YW} w$  und  $\alpha_{YX} = \beta_{YX}$ . Das bedeutet, daß der Steigungskoeffizient  $\alpha_{YX}$  aus Gleichung 4.3.1 invariant bleibt, wenn die potentielle Störvariable W konstant gehalten wird, d.h. W ist keine tatsächliche Störvariable.

Die Invarianz von  $\alpha_{YX}$  ist nicht selbstverständlich. Gilt z.B. weder  $\beta_{YW} = 0$ , noch  $E(W | X) \underset{f_s}{=} E(W)$ , so folgt aus Gleichung 4.3.2 und den oben angegebenen Gleichungen des Anhangs, daß  $\alpha \neq \beta_{YX}$ . Folglich gilt dann bei Konstanz von W nicht mehr der Steigungskoeffizient  $\alpha_{YX}$ . W ist dann tatsächlich eine Störvariable, die den Zusammenhang zwischen X und Y modifiziert. Die Invarianzbedingung postuliert aber genau diese Invarianz des Steigungskoeffizienten bei Konstanz aller potentiellen Störvariablen.

## 4.4 Definition

Nach diesen Überlegungen soll eine stochastische Variable Y von einer zweiten stochastischen Variablen X genau dann einfach kausal reglinear abhängig heißen, wenn erstens X der Variablen Y vorgeordnet ist und zweitens die Invarianzbedingung gilt. Dies soll präziser in der folgenden Definition festgehalten werden:

---

Definition 4.4.1. Es seien X und Y stochastische Variablen auf dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  mit den Erwartungswerten  $E(X)$  bzw.  $E(Y)$ .  $(\mathcal{A}_t, t \in \mathbb{R})$  sei eine isotone Familie von Sigmaalgebren mit  $\& (\bigcup_{t \in \mathbb{R}} \mathcal{A}_t) = \mathcal{A}$ .

Außerdem sei Y von X reglinear abhängig, d.h.

$$(4.4.1) \quad E(Y | X) \underset{f_s}{=} \alpha_{Y0} + \alpha_{YX} X.$$

Die Variable Y heißt von X (bezüglich  $(\mathcal{A}_t, t \in \mathbb{R})$ ) einfach kausal reglinear abhängig und die reelle Zahl  $\alpha_{YX}$  heißt einfacher kausaler reglinearer Effekt von X auf Y genau dann, wenn die folgenden Bedingungen erfüllt sind:

Vorgeordnetheit: X ist Y vorgeordnet<sup>12)</sup>.

---

<sup>12)</sup> Man beachte die Definitionen 4.2.1 und 4.2.2.

*Invarianz:* Für jede potentielle Störvariable W gilt

$$(4.4.2) \quad E(Y|X,W) \stackrel{f_s}{=} \beta_{YX} X + f(W)$$

mit  $\alpha_{YX} = \beta_{YX}$  und  $f(W) = \alpha_{Y0} + E(F|W)$ , wobei  $F := Y - E(Y|X)$ .

Man beachte, daß  $f(W)$  eine Funktion ist, die ausschließlich von W abhängt. Insbesondere wird dadurch eine Interaktion oder Modifizierung der Abhängigkeit der Variablen Y von X durch W ausgeschlossen.

In dieser Definition werden keinerlei formal undefinierte Begriffe verwendet. Es handelt sich also um eine rein mathematische Definition, mit dem damit verbundenen Vorteil, daß eine rein mathematische Untersuchung der daraus ableitbaren Folgerungen möglich ist. Indem man in einem Anwendungsfall festlegt, daß der zugrundegelegte Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  eine ganz bestimmte Situationsklasse oder ein ganz bestimmtes Experiment repräsentiert, daß jede Sigmaalgebra & die bis zum Zeitpunkt t möglichen Ereignisse als Elemente enthält, und daß die Variablen X und Y für ganz bestimmte Variablen stehen, wird eine Interpretation oder Anwendung des formalen Begriffs der einfachen kausalen reglinearen Abhängigkeit vorgenommen.

Ebenso wie beim formalen Begriff der Wahrscheinlichkeit, sind auch hier verschiedene Anwendungen oder Interpretationen denkbar. Zum einen kann man versuchen, mit diesem Begriff objektive kausale reglineare Abhängigkeiten zu beschreiben, zum anderen aber auch *subjektive* Sachverhalte, womit man sich in die psychologische Forschung der Kausalattribution begäbe (vgl. z.B. Heider, 1958, Herkner, 1980, Jones, Kanouse, Kelley, Nisbett, Valins & Weiner, 1971, 1972, und Weiner, 1972, 1974). Sowohl für die erstgenannten methodologischen, als auch für die letztgenannten psychologischen Anwendungen sind die aus der Definition ableitbaren formalen Eigenschaften dieses Begriffs von großer Bedeutung, da sie oft leichter als die in der Definition genannten Eigenschaften empirisch überprüfbar sind. Diesen formalen Eigenschaften und den darauf basierenden empirischen Forschungsstrategien im methodologischen Bereich wenden wir uns im Abschnitt 5 zu.

Zur Erläuterung der Definition 4.4.1 betrachten wir die folgende Situation: Y sei von X einfach kausal reglinear abhängig. Gilt dann beispielsweise außer

$$(4.4.3) \quad E(Y|X) \stackrel{f_s}{=} \alpha_{Y0} + \alpha_{YX} X$$

auch, daß die Residualvariable F von einer potentiellen Störvariablen W reglinear abhängig ist,

$$(4.4.4) \quad E(F|W) \stackrel{f_s}{=} \alpha_{F0} + \alpha_{FW} W,$$

dann muß auch

$$(4.4.5) \quad E(Y | X, W) \stackrel{fs}{=} (\alpha_{Y0} + \alpha_{F0}) + \alpha_{YX} X + \alpha_{FW} W + \alpha_{Y,X \times W} X \cdot W,$$

mit  $\alpha_{Y,X \times W} = 0$

gelten. Zum einen heißt dies, daß der Koeffizient  $\alpha_{YX}$  bei einer Erweiterung des Modells um eine solche Variable  $W$  invariant bleiben muß, und zum anderen, daß keine Interaktion im varianzanalytischen Sinn zwischen  $X$  und  $W$  bestehen darf. Um diese Aussage zu verifizieren, braucht man lediglich die Gleichungen 4.4.4 in Gleichung 4.4.2 einzusetzen:

$$(4.4.6) \quad E(Y | X, W) \stackrel{fs}{=} \alpha_{Y0} + \alpha_{YX} X + \alpha_{F0} + \alpha_{FW} W.$$

Ist in einem Anwendungsfall diese Eigenschaft nicht erfüllt, so können wir schließen, daß  $Y$  von  $X$  nicht einfach kausal reglinear abhängig ist.

Die Definition 4.4.1 hat sowohl für experimentelle, als auch für nichtexperimentelle Forschungssituationen praktische Konsequenzen, da sie für beide Fälle Möglichkeiten aufzeigt, wie man eine Behauptung, daß eine reglineare Abhängigkeit einfach kausal ist, falsifizieren kann. Sowohl das in der Invarianzbedingung enthaltene Interaktionsverbot der Variablen  $X$  mit einer potentiellen Störvariablen  $W$  läßt sich in experimentellen und in nichtexperimentellen Forschungssituationen überprüfen, als auch die darin enthaltene Aussage, daß der Koeffizient  $\alpha_{YX}$  bei der Erweiterung von  $E(Y | X)$  nach  $E(Y | X, W)$  invariant bleibt, d.h. in Gleichung 4.4.1 und in der Gleichung 4.4.2 derselbe ist. Beide Aspekte der Additivitätsbedingung sollen an je einem Beispiel erläutert werden.

## 4.5 Beispiel: Münzen und Elektromagnet (1. Fortsetzung)

Aus Abschnitt 2.3 liegt bereits die Gleichung

$$(4.5.1) \quad E(Z_I | Z_{II}) \stackrel{fs}{=} \beta_{10} + \beta_{12} Z_{II} = 0.5666 + 0.1905 \cdot Z_{II}$$

vor, und wir haben nach Theorem 5.4.1 zu prüfen, ob  $\beta_{12} = 0.1905$  mit dem Koeffizienten  $\alpha_{12}$  aus der Gleichung

$$(4.5.2) \quad E(Z_I | Z_{II}, Z_M) \stackrel{fs}{=} P(B | Z_{II}, Z_M) \stackrel{fs}{=}$$

$$= \alpha_{10} + \alpha_{12} Z_{II} + \alpha_{13} Z_M + \alpha_{14} Z_{II} \cdot Z_M$$

identisch ist, und ob  $\alpha_{14} = 0$  gilt, wobei

$$(4.5.3) \quad B := \{\omega \in \Omega; Z_I(\omega) = 1\}$$

wieder das Ereignis ist, daß Münze I auf die Metallseite fällt.

Die zunächst unbekannten Parameter können wir aus den vier Gleichungen

$$(4.5.4) \quad P(B | Z_{II}=0, Z_M=0) \stackrel{fs}{=} \alpha_{10} = 0.5$$

$$(4.5.5) \quad P(B | Z_{II}=0, Z_M=1) \stackrel{fs}{=} \alpha_{10} + \alpha_{13} = 0.9$$

$$(4.5.6) \quad P(B | Z_{II}=1, Z_M=0) \stackrel{fs}{=} \alpha_{10} + \alpha_{12} = 0.5$$

$$(4.5.7) \quad P(B | Z_{II}=1, Z_M=1) \stackrel{fs}{=} \alpha_{10} + \alpha_{12} + \alpha_{13} + \alpha_{14} = 0.9$$

bestimmen, wobei wir die numerischen Werte für die bedingten Wahrscheinlichkeiten aus den Daten der Tabellen 2.2.1 und 2.2.2 erhalten. Aus obigem Gleichungssystem folgen dann die in der Gleichung

$$(4.5.8) \quad P(B | Z_{II}, Z_M) \stackrel{fs}{=} 0.5 + 0 \cdot Z_{II} + 0.4 \cdot Z_M + 0 \cdot Z_{II} \cdot Z_M$$

angegebenen Koeffizienten. Zwar ist das Interaktionsverbot erfüllt, da  $\alpha_{14} = 0$ , aber es gilt  $\alpha_{12} = 0 \neq \beta_{12} = 0.1905$ , woraus wir nach Definition 4.4.1 schließen können, daß  $Z_I$  nicht von  $Z_{II}$  einfach kausal reglinear abhängig ist.

Damit haben wir ein Beispiel für den ersten in der Invarianzbedingung enthaltenen Aspekt behandelt, der die Gleichheit der Koeffizienten  $\alpha_{12}$  und  $\beta_{12}$  aus den Gleichungen 4.5.1 und 4.5.2 betrifft. Den zweiten Aspekt, nämlich das Interaktionsverbot erläutern wir im folgenden Abschnitt.

## 4.6 Beispiel: Drogen und Aktivierung

In der Tabelle 4.6.1 sind die bedingten Erwartungswerte der abhängigen Variablen

$$(4.6.1) \quad Z_A := \text{„psychophysiologische Aktivierung“}$$

angegeben, die aus einem fiktiven Experiment stammen mögen, in dem die Wirkung der unabhängigen Variablen

$$(4.6.2) \quad Z_D := \begin{cases} 1, & \text{wenn die Versuchsperson das Verum,} \\ -1, & \text{wenn sie das Placebo einnimmt,} \end{cases}$$

untersucht werden soll. Dabei nehmen wir an, daß alle einschlägigen Kontrolltechniken, wie z.B. Randomisierung, Doppelblindversuch u.ä. durchgeführt wurden. Außerdem sei von den Versuchspersonen bekannt, ob sie intro- oder aber extravertiert<sup>13)</sup> sind, was durch die Persönlichkeitsvariable

---

<sup>13)</sup> Die hier vorgenommene Reduzierung der vielstufigen quantitativen Persönlichkeitsvariablen Intro/Extraversion auf zwei Ausprägungen wird ausschließlich aus Gründen der einfacheren Darstellbarkeit des Interaktionsverbots vorgenommen.

$$(4.6.3) \quad Z_p := \begin{cases} 1, & \text{wenn die Versuchsperson introvertiert,} \\ -1, & \text{wenn sie extravertiert ist,} \end{cases}$$

angezeigt wird.

In diesem Beispiel können wir eine eindeutige zeitliche Ordnung zwischen den Variablen vornehmen, denn die Intro/Extraversionsvariable  $Z_p$  ist als Persönlichkeitsvariable zeitlich vor der experimentell manipulierten Drogenvariablen  $Z_D$  anzugeordnen, und letztere zeitlich vor der Aktivierungsvariablen  $Z_A$ , die einen psychophysiologischen Zustand der Versuchspersonen nach der Einnahme der Droge bzw. des Placebos angibt. Falls in diesem Beispiel die Invarianzbedingung erfüllt wäre, könnte die zwischen  $Z_A$  und  $Z_D$  bestehende reglineare Abhängigkeit also kausal interpretiert werden. Ob man in diesem Beispiel diese Voraussetzung machen kann, soll nun überprüft werden.

Tabelle 4.6.1: Bedingte Erwartungswerte der abhängigen Variablen ‚Aktivierung‘ in einem 2x2-kreuzfaktoriellen Design.

	Introvertierte	Extravertierte	
Droge	30	20	25
Placebo	25	5	15
	27.5	12.5	20

Aus den in der Tabelle 4.6.1 angegebenen Zahlen können wir ersehen, daß eine Interaktion im varianzanalytischen Sinn besteht, denn  $30 - 25 \neq 20 - 5$ , d.h. die Differenz der bedingten Erwartungswerte zwischen denjenigen, welche die Droge und denjenigen, die das Placebo einnehmen, ist bei Introvertierten geringer als bei Extravertierten.

Für die bedingte Erwartung  $E(Z_A | Z_D, Z_p)$  gilt

$$(4.6.4) \quad E(Z_A | Z_D, Z_p) = \alpha_0 + \alpha_D Z_D + \alpha_p Z_p + \alpha_{D \times p} Z_D \cdot Z_p = \\ = 20 + 5.0 \cdot Z_D + 7.5 \cdot Z_p + (-2.5) \cdot Z_D \cdot Z_p,$$

wobei wir die numerischen Werte aus den vier Gleichungen

$$(4.6.5) \quad E(Z_A | Z_D=1, Z_p=1) = \alpha_0 + \alpha_D + \alpha_p + \alpha_{D \times p} = 30,$$

$$(4.6.6) \quad E(Z_A | Z_D=1, Z_p=-1) = \alpha_0 + \alpha_D - \alpha_p - \alpha_{D \times p} = 20,$$

$$(4.6.7) \quad E(Z_A | Z_D=-1, Z_p=1) = \alpha_0 - \alpha_D + \alpha_p - \alpha_{D \times p} = 25, \quad \text{und}$$

$$(4.63) \quad E(Z_A | Z_D = -1, Z_P = -1) = \alpha_0 - \alpha_D - \alpha_P + \alpha_{D \times P} = 5$$

berechnen können, die nach den Gleichungen A.5.2, und A.4.6 aus der ersten Formelzeile von 4.6.4 folgen.

Aus Gleichung 4.6.4 können wir ersehen, daß die reglineare Abhängigkeit der Variablen  $Z_A$  von  $Z_D$  nicht einfach kausal ist, denn andernfalls müßte gemäß Definition 4.4.1 der letzte Summand in Gleichung 4.6.4 wegfallen, d.h.  $\alpha_{D \times P}$  dürfte nicht -2.5, sondern müßte gleich Null sein, was dann der Fall wäre, wenn keine Interaktion zwischen der Drogenvariable  $Z_D$  und der Intro/Extraversionsvariablen  $Z_P$  bestände<sup>14)</sup>.

Dieses Beispiel verdeutlicht, daß die Invarianzbedingung einer einfachen kausalen reglinearen Abhängigkeit einer Variablen  $Y$  von  $X$  immer dann nicht gegeben ist, wenn eine Interaktion von  $X$  mit einer potentiellen Störvariablen  $W$ , im Beispiel, der Intro/Extraversionsvariablen, besteht. Ein möglicher Weg wäre nun, die Aussage, daß die Aktivierungsvariable  $Z_A$  einfach kausal reglinear von der Drogenvariablen  $Z_D$  abhängig ist, auf jeweils eine der beiden Gruppen der Intro- bzw. Extravertierten zu beschränken, so daß für jede dieser beiden Gruppen ein unterschiedlicher einfacher kausaler reglinearer Effekt der Drogenvariablen angenommen wird. Die Bedingung der Invarianz könnte dann jeweils innerhalb dieser beiden Gruppen überprüft werden, indem man nach anderen potentiellen Störvariablen sucht, und prüft, ob für diese die in Definition 4.4.1 formulierte Invarianzbedingung erfüllt ist.

Das Interaktionsverbot, das die hier vorgeschlagene Definition einfacher kausaler reglinearer Abhängigkeit beinhaltet, ist zweifellos sehr restriktiv, und es stellt sich die Frage, ob ein derart strenger Begriff einer kausalen reglinearen Abhängigkeit in Wissenschaften wie der Psychologie überhaupt angewandt werden kann. Eine pauschale Antwort auf diese Frage wäre wohl nicht angebracht. Vielleicht müssen wir uns in vielen Bereichen der Psychologie tatsächlich damit zufriedengeben, nur durchschnittliche kausale stochastische Effekte feststellen zu können, im obigen Beispiel möglicherweise also den durchschnittlichen einfachen kausalen reglinearen Effekt der Drogenvariablen  $Z_D$  auf die Aktivierungsvariable  $Z_A$ , als Mittel der Effekte in den beiden Gruppen der Intro- bzw. Extravertierten. Was aber wäre, wenn der durchschnittliche Effekt Null wäre, wohingegen bei den Extravertierten ein starker positiver, bei den Introvertierten dagegen ein ebenso starker negativer Effekt vorläge? Zumindest in Fällen mit einer solchen disordinalen Interaktion (vgl. z.B. Bredenkamp, 1980, S. 24ff. oder Henning & Muthig, 1979, S. 204ff.) kann ein Verzicht, die kausalen reglinearen Abhängigkeiten zu suchen, fatale Konsequenzen haben.

<sup>14)</sup> In realen Forschungssituationen müssen natürlich zur Entscheidung, ob eine Interaktion besteht oder nicht, statistische Verfahren zur Hypothesenbewertung durchgeführt werden, ebenso wie bei der Überprüfung, ob der Koeffizient  $\alpha_{YX}$  (siehe Gleichung 4.4.1) bei der Erweiterung des Modells um  $W$  invariant bleibt (siehe Gleichung 4.4.2). Solche Stichprobenprobleme sind jedoch nicht Gegenstand dieses Beitrags.

## 4.7 Zusammenfassende Bemerkungen

In diesem Abschnitt wurden die im Abschnitt 3 erläuterten intuitiven Vorstellungen interner Validität zum Begriff einfacher kausaler reglinearer Abhängigkeit präzisiert, auf dem die gesamte weitere Theorie aufgebaut ist. Eine solche Abhängigkeit gibt eine reine, unkonfundierte Beziehung an, die durch eine einfache lineare Regressionsgleichung beschrieben werden kann. Es wurden zwei Bedingungen angegeben und erläutert, welche eine einfache kausale regressiv lineare Abhängigkeit einer stochastischen Variablen Y von einer zweiten stochastischen Variablen X definieren: Die Bedingung der Vorgeordnetheit, die in Anwendungsfällen als zeitliche Vorgeordnetheit interpretiert werden kann und die Bedingung der Invarianz, mit der die Existenz von Störvariablen ausgeschlossen wird, die den Zusammenhang zwischen X und Y modifizieren oder verfälschen. Dadurch, daß stochastischen Variablen die Eigenschaft der einfachen kausalen reglinearen Abhängigkeit zugeschrieben wurde, bezieht sich eine solche Aussage immer auf einen ganz bestimmten Wahrscheinlichkeitsraum, der in Anwendungen ein ganz bestimmtes Experiment oder eine ganz bestimmte Untersuchungssituationsklasse repräsentiert. Der Wahrscheinlichkeitsraum ist gewissermaßen der Geltungsbereich einer Aussage über eine einfache kausale reglineare Abhängigkeit, worauf wir ausführlicher im Abschnitt 7 eingehen. Genauso bezieht sich die Vorgeordnetheit auf eine ganz bestimmte isotone Familie ( $\mathcal{A}_t, t \in \mathbb{R}$ ) von Sigmaalgebren, beispielsweise auf die in einem ganz bestimmten Experiment gegebene Familie von Sigmaalgebren  $\mathcal{A}_t$ , welche die bis zum Zeitpunkt t möglichen Ereignisse enthalten.

## 5. Eigenschaften einfacher kausaler reglinearer Abhängigkeit

### 5.1 Einleitende Bemerkungen

In diesem Abschnitt untersuchen wir die formalen Eigenschaften des im letzten Abschnitt definierten Begriffs der einfachen kausalen reglinearen Abhängigkeit und behandeln die sich daraus ergebenden Konsequenzen für die Strategien empirischer Untersuchungen. Dabei sind zwei Fragen von wesentlicher Bedeutung: Wie überprüft und falsifiziert man in einer vorliegenden empirischen Untersuchung die Aussage, daß eine einfache kausale reglineare Abhängigkeit besteht, und zweitens, wie kann man eine noch durchzuführende Untersuchung so planen und anlegen, daß die Annahme, daß eine einfache kausale reglineare Abhängigkeit vorliegt, möglichst allen Falsifizierungsversuchen standhält.

Im Abschnitt „Unkonfundiertheit“ zeigen wir zunächst, daß der einfache kausale reglineare Effekt  $\alpha_{YX}$  von X auf Y auch bei Konstanz potentieller Störva-

riablen gilt, falls Y von X einfach kausal reglinear abhängig ist. Am Beispiel ‚Münzen und Elektromagnet‘ demonstrieren wir dann, wie diese Eigenschaft benutzt werden kann, um die Aussage, daß eine einfache kausale reglineare Abhängigkeit der ersten von der zweiten Münzvariablen besteht, zu falsifizieren.

Im Abschnitt ‚vollständige Abhängigkeit‘ zeigen wir, welcher Art der deterministische Kausalitätsbegriff ist, der in der hier vorgestellten Theorie als Spezialfall enthalten ist. Im Abschnitt ‚faktische Konstanthaltung‘ behandeln wir dann eine erste Maßnahme zur Planung von Untersuchungen, mit dem Ziel, die Validität einer reglinearen Abhängigkeit herzustellen. Schließlich formulieren wir im Abschnitt ‚Unabhängigkeit‘ eine notwendige Bedingung einfacher kausaler reglinearer Abhängigkeit und zeigen, wie die experimentellen Kontrolltechniken der Randomisierung und der Parallelisierung auf die hier vorgestellte Theorie basiert werden können.

## 5.2 Unkonfundiertheit

Wir zeigen als erstes, daß  $\alpha_{YX}$  auch der Steigungskoeffizient der einfachen linearen Regression der Variablen Y unter X bei Konstanz einer potentiellen Störvariablen ist, wenn wir voraussetzen können, daß Y von X einfach kausal reglinear abhängig ist. Daraus ergibt sich eine Möglichkeit zur Falsifizierung einer Aussage, daß Y von X einfach kausal reglinear abhängig ist, da sich leicht überprüfen läßt, ob die genannten Steigungskoeffizienten tatsächlich identisch sind oder nicht.

**Theorem 5.2.1.** Es mögen wieder die Voraussetzungen und Schreibweisen gelten, die in der Definition der einfachen kausalen reglinearen Abhängigkeit aufgeführt sind.

Wenn Y von X einfach kausal reglinear abhängig ist, mit dem einfachen kausalen reglinearen Effekt  $\alpha_{YX}$  von X auf Y, dann gilt außer

$$(5.2.1) \quad E(Y | X) \underset{fs}{=} \alpha_{Y0} + \alpha_{YX} X,$$

auch für jedes Ereignis

$$(5.2.2) \quad A := \{\omega \in \Omega: W(\omega) = w\}, \text{ mit } P(A) > 0,$$

daß eine potentielle Störvariable W einen bestimmten Wert w annimmt

$$(5.2.3) \quad E_A(Y | X) \underset{P_A-fs}{=} \alpha_{Y0} + E(F | A) + \alpha_{YX} X,$$

wobei

$$(5.2.4) \quad F :=_{fs} Y - E(Y | X)$$

und  $E(F | A)$  der bedingte Erwartungswert von  $F$  gegeben das Ereignis  $A$  ist.

---

Beweis. Nach den Definitionen 4.4.1 und A.5.1 gilt

$$(5.2.5) \quad E(1_D Y) = E[1_D [\alpha_{Y0} + \alpha_{YX} X + E(F | W)]],$$

für alle  $D \in \mathcal{A}(X, W)$ .

Also gilt auch für alle  $A$  mit  $P(A) > 0$ , die durch 5.2.2 definiert sind, und alle  $C \in \mathcal{A}(X)$

$$(5.2.6) \quad \frac{1}{P(A)} E(1_A 1_C Y) = \frac{1}{P(A)} E[1_A 1_C [\alpha_{Y0} + \alpha_{YX} X + E(F | W)]].$$

Daraus folgt nach Gleichung A.4.1

$$(5.2.7) \quad E_A(1_C Y) = E_A[1_C [\alpha_{Y0} + E(F | W) + \alpha_{YX} X]] =$$

$$= E_A[1_C [\alpha_{Y0} + E(F | A) + \alpha_{YX} X]],$$

denn  $E_A[1_C E(F | W)] = E_A[1_C E(F | A)]$ , da  $E(F | W)$  bezüglich des Maßes  $P_A$  eine Konstante ist. Nach Definition A.5.1 folgt dann Gleichung 5.2.3, denn die Meßbarkeit von  $E_A(Y | X)$  bezüglich  $\mathcal{A}(X)$  ist offensichtlich erfüllt.

Theorem 5.2.1 zeigt, daß die Definition der einfachen kausalen reglinearen Abhängigkeit die Eigenschaft beinhaltet, daß potentielle Störvariablen die Beziehung zwischen  $X$  und  $Y$  nicht modifizieren oder verfälschen und in diesem Sinne stören, falls  $Y$  von  $X$  einfach kausal reglinear abhängig ist. Der gleiche Steigungsparameter  $\alpha_{YX}$  gilt diesem Theorem zufolge nämlich unabhängig davon, ob solche Variablen konstant sind oder nicht. Mit anderen Worten, die Abhängigkeit der Variablen  $Y$  von  $X$  ist nicht durch andere Variablen konfundiert, wenn  $Y$  von  $X$  einfach kausal reglinear abhängig ist. Der Steigungskoeffizient  $\alpha_{YX}$  aus Gleichung 5.2.1 charakterisiert dann also eine reine, unkonfundierte reglineare Abhängigkeit der Variablen  $Y$  von  $X$ . Die in Theorem 5.2.1 formulierte Eigenschaft nennen wir daher Unkonfundiertheitsbedingung einfacher kausaler reglinearer Abhängigkeit.

### 5.3 Beispiel: Münzen und Elektromagnet (1. Fortsetzung)

Wir wenden nun Theorem 5.2.1 auf die in Abschnitt 2.3 beschriebene reglineare Abhängigkeit der ersten von der zweiten Münzvariablen an. In dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$ , der die in Abschnitt 2.2 beschriebene dritte Situationsklasse (siehe Tabelle 2.2.3) repräsentiert, gilt zwar für die beiden Münzvariablen  $Z_I$  und  $Z_{II}$

$$(53.1) \quad E(Z_I | Z_{II}) \stackrel{fs}{=} \beta_{10} + \beta_{12} Z_{II} = 0.5666 + 0.1905 \cdot Z_{II}$$

(siehe die Gleichungen 2.3.4 bis 2.3.7), und  $Z_{II}$  ist  $Z_I$  vorgeordnet, nämlich nach dem Gesichtspunkt der Handlungsabfolge, aber die durch Gleichung 5.3.1 beschriebene reglineare Abhangigkeit ist nicht einfach kausal, da die Unkonfundiertheitsbedingung nicht erfullt ist.

Um dies zu zeigen, betrachten wir die stochastische Variable

$$(5.3.2) \quad Z_M = \begin{cases} 1, & \text{wenn der Elektromagnet an,} \\ 0, & \text{wenn er aus ist,} \end{cases}$$

die der Variablen  $Z_{II}$  zeitlich vorgeordnet ist (siehe Abbildung 4.2.1). Gema der im Theorem 5.2.1 formulierten notwendigen Bedingung konnte die durch Gleichung 5.3.1 beschriebene reglineare Abhangigkeit nur dann einfach kausal sein, wenn z.B. fur das Ereignis

$$(5.3.3) \quad A := \{\omega \in \Omega : Z_M(\omega) = 1\},$$

da der Elektromagnet angeschaltet ist, die Gleichung

$$(5.3.4) \quad E_A(Z_I | Z_{II}) \stackrel{P_A, fs}{=} \alpha_{10} + \alpha_{12} Z_{II},$$

mit  $\alpha_{12} = 0.1905$  galte, wenn also auch fur die einfache lineare Regression der Variablen Y unter X gegeben A, der Steigungskoeffizient 0.1905 gelten wurde. Die analoge Aussage wurde naturlich fur das Ereignis

$$(5.3.5) \quad \bar{A} := \{\omega \in \Omega : Z_M(\omega) = 0\}$$

zutreffen.

Zur Berechnung des tatsachlich gultigen Koeffizienten  $\alpha_{12}$  in Gleichung 5.3.4 bilden wir fur das Ereignis

$$(5.3.6) \quad B := \{\omega \in \Omega : Z_I(\omega) = 1\}.$$

da die Munze I auf die Metallseite fallt, nach den Gleichungen A.5.2, A.4.3 und A.4.6 des Anhangs die beiden bedingten Wahrscheinlichkeiten bezuglich des Maes  $P_A$  von B gegeben  $Z_{II} = 0$  bzw.  $Z_{II} = 1$ .

$$(5.3.7) \quad E_A(Z_I | Z_{II}=0) = P_A(B | Z_{II}=0) = \alpha_{10}$$

und

$$(5.3.8) \quad E_A(Z_I | Z_{II}=1) = P_A(B | Z_{II}=1) = \alpha_{10} + \alpha_{12}.$$

Bezeichnen wir mit

$$(5.3.9) \quad C := \{\omega \in \Omega : Z_{II}(\omega) = 1\}$$

und

$$(5.3.10) \quad \bar{C} := \{\omega \in \Omega: Z_{II}(\omega) = 0\}$$

die Ereignisse, daß die Münze II auf die Metall- bzw. Plastikseite fällt, so erhalten wir aus den Daten der Tabelle 2.2.2 die folgenden bedingten Wahrscheinlichkeiten:

$$(5.3.11) \quad P_A(B | Z_{II}=0) = P(B | Z_{II}=0, Z_M=1) = \\ = \frac{P(B \cap \bar{C} | Z_M=1)}{P(\bar{C} | Z_M=0)} = \frac{0.09 \cdot 0.50}{0.10 \cdot 0.50} = \frac{0.09}{0.10} = 0.9$$

und

$$(5.3.12) \quad P_A(B | Z_{II}=1) = P(B | Z_{II}=1, Z_M=1) = \\ = \frac{P(B \cap C | Z_M=1)}{P(C | Z_M=1)} = \frac{0.81 \cdot 0.50}{0.90 \cdot 0.50} = \frac{0.81}{0.90} = 0.9.$$

Aus den Gleichungen 5.3.7 und 5.3.8 folgt dann

$$(5.3.13) \quad \alpha_{10} = 0.9 \quad \text{und} \quad \alpha_{10} + \alpha_{12} = 0.9,$$

woraus sich  $\alpha_{12} = 0$  ergibt.

Aus  $\alpha_{12} = 0$  können wir aber schließen, daß die Bedingung der Unkonfundiertheit für die reglineare Abhängigkeit der Variablen  $Z_I$  von der Variablen  $Z_{II}$  nicht erfüllt ist, denn  $\alpha_{12} = 0 \neq \beta_{12} = 0.1905$ , d.h. die Steigungskoeffizienten der Gleichungen 5.3.1 und 5.3.4 sind nicht identisch, wie es die Unkonfundiertheitsbedingung (siehe Theorem 5.2.1) fordert. Daraus folgt aber, daß die reglineare Abhängigkeit der ersten Münzvariablen  $Z_I$  von der zweiten Münzvariablen  $Z_{II}$  im Beispiel „Münzen und Elektromagnet“ nicht einfach kausal ist, ein Schluß, der mit unserer intuitiven Vorstellung wohl voll im Einklang steht. Der hier definierte Kausalitätsbegriff ist demnach geeignet, eine kausale Interpretation der Abhängigkeit der ersten von der zweiten Münzvariablen im Beispiel „Münzen und Elektromagnet“ auszuschließen. Damit ist eine weitere Möglichkeit aufgezeigt, wie man überprüfen kann, ob eine reglineare Abhängigkeit einfach kausal ist.

## 5.4 Vollständige Abhängigkeit

Bevor wir uns den experimentellen Kontrolltechniken zuwenden, sei eine erste hinreichende Bedingung einfacher kausaler reglinearer Abhängigkeit genannt, die besagt, daß eine solche Abhängigkeit gegeben ist, wenn außer der Vorgeordnetheitsbedingung auch gilt, daß  $Y$  vollständig von  $X$  abhängig ist, derart, daß  $Y \stackrel{fs}{=} E(Y | X)$ . Damit betrachten wir den Fall, den man umgangssprach-

lich als ‚deterministisch‘ bezeichnet<sup>15)</sup>). Dabei darf jedoch nicht außer acht gelassen werden, daß es sich auch hier um eine stochastische Abhängigkeit stochastischer Variablen handelt. Stochastische Abhängigkeiten sind also in dem hier verwendeten Sprachgebrauch (vgl. z.B. Bauer, 1974, S. 137 und S. 150) nicht unbedingt unvollkommen oder unvollständig.

---

Theorem 5.4.1. Es mögen wieder die Voraussetzungen und Schreibweisen gelten, die in der Definition der einfachen kausalen reglinearen Abhängigkeit aufgeführt sind.

Wenn die Bedingung der Vorgeordnetheit erfüllt ist, und es gilt

$$(5.4.1) \quad Y \underset{\text{fs}}{=} E(Y | X) \underset{\text{fs}}{=} \alpha_{YO} + \alpha_{YX} X,$$

dann ist  $Y$  von  $X$  einfach kausal reglinear abhängig mit dem einfachen kausalen reglinearen Effekt  $\alpha_{YX}$ .

---

*Beweis.* Aus Gleichung 5.4.1 folgt für jede stochastische Variable  $W$  auf  $(\Omega, \mathcal{A}, P)$

$$(5.4.2) \quad Y \underset{\text{fs}}{=} E(Y | X) \underset{\text{fs}}{=} \alpha_{YO} + \alpha_{YX} X \underset{\text{fs}}{=} E(Y | X, W),$$

und  $E(F | W) \underset{\text{fs}}{=} 0$ , da bereits  $F \underset{\text{fs}}{=} Y - E(Y | X) \underset{\text{fs}}{=} 0$  ist. Daher gilt aber auch die Gleichung 4.4.2 für jede potentielle Störvariable  $W$ , was zu beweisen war.

Im Fall vollständiger Abhängigkeit, in der die Gleichung 5.4.1 erfüllt ist, muß lediglich dafür gesorgt werden, daß die Bedingung der Vorgeordnetheit erfüllt ist.

Bei vollständiger Abhängigkeit lassen sich Beispiele konstruieren, die zunächst nicht mit unserer Intuition übereinzustimmen scheinen. Man stelle sich ein Experiment ähnlich wie das in Abschnitt 2 beschriebene vor, jedoch mit zwei Münzen I und II, in die jeweils auf einer Seite ein Dauermagnet eingebaut ist. Auch hier wird zuerst Münze II, dann Münze I geworfen, wieder auf eine Platte, welche die Eigenschaften eines Elektromagneten besitzt. Dieser soll nun mit Wahrscheinlichkeit 0.5 in einem von zwei Zuständen sein: In der einen Richtung gepolt, sei der Elektromagnet so stark, daß beide Münzen mit Wahrscheinlichkeit 1 auf die Seite A fallen, und in der anderen Richtung gepolt, sei er ebenfalls so stark, daß die beiden Münzen mit Wahrscheinlichkeit 1 auf die Seite B fallen. Die beiden Münzvariablen

---

<sup>15)</sup> Vgl. zu deterministischen Kausalbegriffen z.B. Popper (1969, S. 31f.) und Essler (1979).

$$(5.4.3) \quad Z_I := \begin{cases} 1, & \text{wenn Münze I auf Seite A,} \\ 0, & \text{wenn sie auf Seite B fällt} \end{cases}$$

und

$$(5.4.4) \quad Z_{II} := \begin{cases} 1, & \text{wenn Münze II auf Seite A,} \\ 0, & \text{wenn sie auf Seite B fällt,} \end{cases}$$

sind dann vollständig abhängig von der Magnetvariablen

$$(5.4.5) \quad Z_M := \begin{cases} 1, & \text{wenn der Elektromagnet in der einen,} \\ 0, & \text{wenn er in der anderen Richtung gepolt ist.} \end{cases}$$

Dadurch bedingt ist auch die erste Münzvariable  $Z_I$  vollständig von der zweiten Münzvariablen  $Z_{II}$  abhängig.

Demnach gälte dann

$$(5.4.6) \quad Z_I \underset{\text{fs}}{=} E(Z_I | Z_{II}) \underset{\text{fs}}{=} 0 + 1 \cdot Z_{II}$$

und da die Münze II zuerst geworfen wird,  $Z_{II}$  also  $Z_I$  vorgeordnet ist, gilt dann, daß  $Z_I$  von  $Z_{II}$  einfach kausal reglinear abhängig ist, eine Aussage, die zunächst im Widerspruch zu unserer Intuition zu stehen scheint. Wenn man allerdings bedenkt, daß sich eine Aussage über eine einfache kausale reglineare Abhängigkeit auf einen ganz bestimmten Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  bezieht, so schwächt sich dieser Eindruck ab. In dem hier zugrundegelegten Wahrscheinlichkeitsraum ist nämlich tatsächlich immer  $Z_I = 1$ , wenn  $Z_{II} = 1$  und  $Z_I = 0$ , wenn  $Z_{II} = 0$  ist. Der zugrunde gelegte Wahrscheinlichkeitsraum ist der Geltungsbereich einer Aussage über stochastische und daher auch über kausale stochastische Abhängigkeiten, worauf wir ausführlicher in den Abschnitten 6 und 7 zurückkommen.

Der Nachteil der Aussage, daß  $Z_I$  von  $Z_{II}$  in dem hier zugrundegelegten Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  mit dem Effekt 1.0 einfach kausal reglinear abhängig ist, besteht nicht darin, daß sie falsch ist, sondern darin, daß der Geltungsbereich extrem eng ist, d.h. daß der Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  extrem „klein“ ist, denn diese Aussage gilt ja nur in der oben beschriebenen Situationsklasse, bei welcher der Elektromagnet nur die beiden Zustände „stark positiv gepolt“ und „stark negativ gepolt“ annehmen kann. Darüber hinaus kann innerhalb dieser Situationsklasse nicht zwischen den Aussagen „ $Z_I$  ist von  $Z_{II}$ “ bzw. „ $Z_I$  ist von  $Z_M$  einfach kausal reglinear abhängig“ entschieden werden, d.h. hier liegt ein Problem der Identifizierbarkeit vor (vgl. z.B. Anderson & Rubin, 1956; Fischer, 1966, 1978, und Wiley, 1973). Würden wir den Wahrscheinlichkeitsraum erweitern, so daß der Elektromagnet auch den Zustand „aus“ annehmen kann, so gälte in diesem größeren Wahrscheinlichkeitsraum nicht mehr die Aussage, daß  $Z_I$  von  $Z_{II}$  mit dem Effekt 1 einfach

kausal reglinear abhängig ist. In diesem größeren Wahrscheinlichkeitsraum würde dann nur noch die allgemeinere Aussage gelten, daß die Münzvariablen  $Z_I$  und  $Z_{II}$  beide von der Magnetvariablen  $Z_M$  kausal stochastisch abhängig sind<sup>16)</sup>.

Der Wahrscheinlichkeitsraum ist ein unverzichtbarer Bestandteil jeder Aussage über stochastische, auch über kausale stochastische Abhängigkeiten, eine Tatsache, die allzu oft in der psychologischen Literatur vernachlässigt wird, wenn beispielsweise davon die Rede ist, daß zwei Variablen in einem bestimmten Ausmaß miteinander korreliert sind. Eine solche Aussage hat nur einen Sinn, wenn die Menge der Situationen und der Individuen (und damit der Wahrscheinlichkeitsraum), für die diese Aussage gelten soll, wenigstens implizit mit angegeben wird.

## 5.5 Faktische Konstanthaltung

Die im Beispiel ‚Drogen und Aktivierung‘ bereits angesprochene Einschränkung des Geltungsbereichs auf jeweils eine der beiden Gruppen der Intro- bzw. Extravertierte, würde man formal dadurch ausdrücken, daß jeweils ein verschiedener Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P_1)$  bzw.  $(\Omega, \mathcal{A}, P_2)$  zugrunde gelegt wird. In dem einen beträgt dann die Wahrscheinlichkeit, daß eine Person introvertiert ist, Eins, und daß sie extravertiert ist, Null. In dem anderen Wahrscheinlichkeitsraum verhält es sich genau umgekehrt. Dies kann man auch als ‚faktische Konstanthaltung‘ der Störvariablen Intro/Extraversion auffassen. Dabei benutzen wir das Wort ‚faktisch‘, um diese Art der Konstanthaltung von der ‚statistischen Konstanthaltung‘ abzugrenzen, auf die wir später noch eingehen werden.

Die Kontrolltechnik der faktischen Konstanthaltung potentieller Störvariablen hat zwei operationale Aspekte, den der *Selektion von Beobachtungseinheiten* und den der *Herstellung gleicher Bedingungen*. Bei der Selektion werden z.B. nur introvertierte oder nur extravertierte, nur männliche oder nur weibliche Versuchspersonen herangezogen etc., womit die Intro/Extraversionsvariable bzw. die Geschlechtsvariable konstant gehalten wird. Bei der Herstellung gleicher Bedingungen handelt es sich z.B. darum, in einer Versuchs- und einer Kontrollgruppe gleiche Beleuchtungsbedingungen herzustellen, so daß die Beleuchtungsvariable konstant gehalten wird. Bei der Selektion handelt es sich also darum, bestimmte Versuchspersonen auszuwählen, und bei der Herstellung gleicher Bedingungen werden bestimmte situative Bedingungen ausgewählt und konstant gehalten.

---

<sup>16)</sup> Die Abhängigkeiten der Münzvariablen  $Z_I$  und  $Z_{II}$  von der Magnetvariablen  $Z_M$  ließen sich dann übrigens auch nicht mehr mit einem reglinearen Modell beschreiben, da der Zusammenhang dann nicht mehr linear im Sinne einer Geradengleichung wäre (siehe dazu auch die Abschnitte 6 und 7).

Mit der Konstanthaltung einer potentiellen Störvariablen  $W$  ist garantiert, daß die Gleichung 4.4.2 für die betreffende Variable  $W$  erfüllt ist, wobei allerdings zu beachten ist, daß mit der Konstanthaltung die betreffende Variable  $W$  zur Konstanten degeneriert ist.

Der zweifellos sicherste Weg, eine empirische Untersuchung so anzulegen, daß die Invarianzbedingung erfüllt ist, bestände darin, dafür zu sorgen, daß alle potentiellen Störvariablen, jeweils auf einem bestimmten Wert konstant sind. In einem solchen Fall wäre dann für jede potentielle Störvariable  $W$  die Gleichung 4.4.2 erfüllt. Diese hinreichende Bedingung für eine einfache kausale reglineare Abhängigkeit soll im folgenden Theorem festgehalten werden.

**Theorem 5.5.1.** Es mögen wieder die Voraussetzungen und Schreibweisen gelten, die in der Definition der einfachen kausalen reglinearen Abhängigkeit aufgeführt sind.

Wenn die Bedingung der Vorgeordnetheit<sup>17)</sup> erfüllt ist, alle potentiellen Störvariablen konstant sind, und die Gleichung

$$(5.5.1) \quad E(Y|X) \stackrel{fs}{=} \alpha_{Y0} + \alpha_{YX} X$$

gilt, dann ist  $Y$  von  $X$  einfach kausal reglinear abhängig mit dem einfachen kausalen reglinearen Effekt  $\alpha_{YX}$ .

**Beweis.** Wenn alle potentiellen Störvariablen  $W$  konstant sind, so ist für alle  $W$  auch die Gleichung 4.4.2 erfüllt, denn  $E(Y|X) \stackrel{fs}{=} E(Y|X,W)$  und  $E(F|W) \stackrel{fs}{=} 0$ , wenn  $W$  eine Konstante ist, womit die Behauptung bewiesen ist.

Die auf diesem Theorem beruhende Methode der Konstanthaltung aller potentiellen Störvariablen, die wohl auf die „Unterschiedsmethode“ von Mill (siehe z.B. 1885, S. 91) zurückgeht, ist jedoch fast nie anwendbar. Selbst bei physikalischen Experimenten können nicht alle potentiellen Störvariablen konstant gehalten werden, denn jedes Ereignis  $A \in \mathcal{A}$ , welches der experimentell manipulierten Variablen  $X$  vorgeordnet ist, kann über die Indikatorvariable  $1_A$  als potentielle Störvariable  $W$  angesehen werden.

In gewisser Hinsicht ist dieses Theorem allerdings von Bedeutung, denn gälte es nicht, dann wäre die Definition der einfachen kausalen reglinearen Abhängigkeit wohl unbrauchbar. Variiert nur die Variable  $X$ , während alle potentiellen Störvariablen konstant sind, dann muß die Abhängigkeit zwischen  $X$  und  $Y$  kausal interpretiert werden können, falls  $X$  der Variablen  $Y$  zeitlich vorgeordnet ist. Ein anderes Ergebnis wäre mit unserer Intuition wohl kaum zu vereinbaren (vgl. z.B. auch McGuigan, 1978, S. 150).

<sup>17)</sup> Siehe Definition 4.4.1.

Die Bedingung der Konstanz aller potentiellen Störvariablen W ist zwar hinreichend, jedoch keineswegs notwendig dafür, daß eine einfache kausale reglineare Abhängigkeit der Variablen Y von X besteht. Wenn nämlich die Bedingung der Vorgeordnetheit vorausgesetzt wird, und außerdem gilt, daß nur solche potentiellen Störvariablen W nicht konstant sind, für die gilt, daß sie die Gleichung 4.4.2 erfüllen, dann ist nach Definition 4.4.1 die Invarianzbedingung der reglinearen Abhängigkeit der Variablen Y von X ebenfalls gegeben, denn für eine konstante stochastische Variable W ist die Gleichung 4.4.2 trivialerweise erfüllt.

Die hierauf basierende experimentelle Kontrolltechnik der Konstanthaltung aller relevanten potentiellen Störvariablen, d.h. aller potentiellen Störvariablen W, für die andernfalls die Gleichung 4.4.2 nicht erfüllt wäre, ist wohl eine prinzipiell gangbare Forschungsstrategie.

Bei dieser Kontrolltechnik ist jedoch zu beachten, daß mit der Konstanthaltung auch die entsprechende Einschränkung des Geltungsbereichs der Aussage über eine stochastische Abhängigkeit verbunden ist, nämlich die Einschränkung auf die betreffende Konstanzbedingung. Wird beim Beispiel ‚Drogen und Aktivierung‘ die Intro/Extraversionsvariable konstant gehalten, so bezieht sich dann die Aussage über eine einfache kausale reglineare Abhängigkeit nur auf jeweils eine der beiden Gruppen der Intro- bzw. Extravertierten. Auf diese mit der Konstanthaltung verbundene Einschränkung des Geltungsbereichs werden wir im Abschnitt 7 ausführlicher eingehen.

## 5.6 Unabhängigkeit

Die im Abschnitt 5.5 behandelten Kontrolltechniken der faktischen Konstanthaltung sind nicht die einzigen Mittel, die uns zur Verfügung stehen, um eine Untersuchung so anzulegen, daß nach Möglichkeit die beobachtete stochastische Abhängigkeit der kausalen entspricht. Andere Kontrolltechniken sind z.B. die Randomisierung und die Parallelisierung. In diesem Abschnitt behandeln wir ein Theorem, auf das diese Kontrolltechniken empirischer Untersuchungen basiert werden können.

Theorem 5.6.1. Es mögen wieder die Voraussetzungen und Schreibweisen gelten, die in der Definition der einfachen kausalen reglinearen Abhängigkeit aufgeführt sind.

Wenn Y von X einfach kausal reglinear abhängig ist, dann gilt für jede potentielle Störvariable W die Gleichung

$$(5.6.1) \quad E[E(F|W)|X] =_s 0, \quad \text{wobei } F :=_s Y - E(Y|X).$$

Beweis. Wenn Y von X einfach kausal reglinear abhängig ist, dann gilt auch die Gleichung 4.4.2, aus der wir wegen der Gleichungen A.5.7, A.5.10 und A.5.5

$$\begin{aligned}
 (5.6.2) \quad E[E(Y | X, W) | X] &\stackrel{\text{fs}}{=} E(Y | X) \stackrel{\text{fs}}{=} \\
 &\stackrel{\text{fs}}{=} E[\alpha_{Y0} + \alpha_{YX} X | X] + E[E(F | W) | X] \stackrel{\text{fs}}{=} \\
 &\stackrel{\text{fs}}{=} \alpha_{Y0} + \alpha_{YX} X + E[E(F | W) | X],
 \end{aligned}$$

erhalten, die nur dann keinen Widerspruch beinhaltet, wenn Gleichung 5.6.1 erfüllt ist, da  $E(Y | X) \stackrel{\text{fs}}{=} \alpha_{Y0} + \alpha_{YX} X$ , falls Y von X einfach kausal reglinear abhängig ist.

Zum besseren Verständnis dieses Theorems überlegen wir zunächst, in welchen Fällen die Gleichung 5.6.1 für eine potentielle Störvariable erfüllt ist. Diese Gleichung gilt erstens dann, wenn die Residualvariable F gleich Null ist, wenn also  $Y = \alpha_{Y0} + \alpha_{YX} X$ , d.h. wenn Y vollständig von X abhängig ist (siehe Abschnitt 5.5). Sie ist zweitens aber auch in dem Fall gegeben, wenn F und W stochastisch unabhängig sind, oder wenn zumindest

$$(5.6.3) \quad E(F | W) \stackrel{\text{fs}}{=} E(F) = 0$$

gilt, woraus nach Theorem A.5.4 folgt, daß F und W auch unkorreliert sind. Diese Situation ist in Abbildung 3.4.2 dargestellt. Schließlich ist Gleichung 5.6.1 drittens auch dann erfüllt, wenn W und X stochastisch unabhängig sind, denn dann gilt wegen der  $\mathcal{A}(W)$ -Meßbarkeit<sup>18)</sup> von  $E(F | W)$ , daß auch  $E(F | W)$  und X stochastisch unabhängig sind, woraus nach Theorem A.5.5 und der Gleichung

$$(5.6.4) \quad E[E(F | W)] = E(F) = 0$$

ebenfalls die Gleichung 5.6.1 folgt. Aus diesem Grund bezeichnen wir die in Theorem 5.6.1 formulierte Eigenschaft als „Unabhängigkeitsbedingung“.

Diese Situation, die man in Experimenten mit den Kontrolltechniken der Randomisierung und der Parallelisierung herstellt, ist in Abbildung 3.4.1 skizziert. Man beachte jedoch, daß die in Theorem 5.6.1 formulierte Unabhängigkeitsbedingung schwächer als die stochastische Unabhängigkeit zwischen W und X, und auch schwächer als die stochastische Unabhängigkeit zwischen  $E(F | W)$  und X ist.

<sup>18)</sup> Eine stochastische Variable Z [im obigen Fall:  $E(E | W)$ ] heißt  $\mathcal{B}$ -meßbar oder meßbar bezüglich der Sigmaalgebra  $\mathcal{B}$  [im obigen Fall: bezüglich  $\mathcal{A}(W)$ ] genau dann, wenn die von Z erzeugte Sigmaalgebra  $\mathcal{A}(Z)$  Teilmenge von  $\mathcal{B}$  ist (vgl. z.B. Müller, 1975, S. 164f.) Die bedingte Erwartung ist so definiert, daß  $E(E | W)$  meßbar bezüglich  $\mathcal{A}(W)$  ist (siehe Definition A.5.1) und aus der stochastischen Unabhängigkeit von W und X folgt, der Definition der stochastischen Unabhängigkeit (siehe z.B. Bauer, 1974, S. 150) gemäß, daß auch  $E(E | W)$  und X stochastisch unabhängig sind.

Bevor wir ausführlicher auf die Randomisierung eingehen, betrachten wir noch die nachstehende Folgerung aus Theorem 5.6.1, die für die Falsifikation einer Aussage, daß Y von X einfach kausal reglinear abhängig ist, verwendet werden kann.

---

Theorem 5.6.2. Es mögen wieder die Voraussetzungen und Schreibweisen gelten, die in der Definition der einfachen kausalen reglinearen Abhängigkeit aufgeführt sind.

Wenn Y von X einfach kausal reglinear abhängig ist mit

$$(5.65) \quad E(Y|X) \stackrel{fs}{=} \alpha_{Y0} + \alpha_{YX} X,$$

und für eine potentielle Störvariable W gilt

$$(5.66) \quad E(F|W) \stackrel{fs}{=} \alpha_{F0} + \alpha_{FW} W, \quad \text{wobei } F := Y - E(Y|X),$$

dann folgt

$$(5.67) \quad E(Y|X,W) \stackrel{fs}{=} (\alpha_{Y0} + \alpha_{F0}) + \alpha_{YX} X + \alpha_{FW} W$$

und wenn außerdem  $\alpha_{FW} \neq 0$ , dann gilt auch

$$(5.68) \quad E(W|X) \stackrel{fs}{=} E(W),$$

und daß X und W unkorreliert sind.

---

*Beweis.* Gleichung 5.6.7 folgt direkt aus Definition 4.4.1 und Gleichung 5.6.6. Nach den Gleichungen A.5.7, A.5.10 und A.5.5 erhalten wir aus Gleichung 5.6.7

$$(5.69) \quad E[E(Y|X,W)|X] \stackrel{fs}{=} E(Y|X) \stackrel{fs}{=} \alpha_{Y0} + \alpha_{YX} X + \alpha_{F0} + \alpha_{FW} E(W|X).$$

Da voraussetzungsgemäß Gleichung 5.6.5 gilt, muß auch  $\alpha_{F0} = \alpha_{FW} E(W|X)$  gelten, was unter der Voraussetzung  $\alpha_{FW} \neq 0$  nur dann zutreffen kann, wenn  $E(W|X) = E(W)$ . Dies aber impliziert nach Theorem A.3.4, daß X und W unkorreliert sind, womit die Behauptung bewiesen ist.

Aus  $\alpha_{FW} \neq 0$  und einer Korrelation von X und W ungleich Null, können wir also schließen, daß Y von X nicht einfach kausal reglinear abhängig ist, wenn wir Gleichung 5.6.6 voraussetzen. Gelten dagegen Gleichung 5.6.7, derzufolge keine Interaktion zwischen X und W besteht, und Gleichung 5.6.8, welche die Unkorreliertheit von X und W impliziert, so modifiziert W nicht die Abhängigkeit der Variablen Y von X, d.h. durch die Variable W wird dann nicht verhindert, daß die Koeffizienten  $\alpha_{YX}$  in den Gleichungen 5.6.5 und 5.6.7 identisch sind, daß also möglicherweise die reglineare Abhängigkeit der Variablen Y von X einfach kausal ist. Eine Anwendung dieses Prinzips besprechen wir u.a. im folgenden Abschnitt.

## 5.7 Randomisierung und Parallelisierung

Die in Gleichung 5.6.1 formulierte Unabhängigkeitsbedingung einfacher kausaler reglinearer Abhängigkeit kann für viele potentielle Störvariablen W z.B. dadurch erfüllt werden, daß man bei einer empirischen Untersuchung die Kontrolltechnik der Randomisierung anwendet, bei der die Beobachtungseinheiten (Versuchspersonen, -tiere, -gruppen etc.) zufällig auf die verschiedenen experimentellen Bedingungen aufgeteilt werden, so daß zwischen den experimentellen Gruppen höchstens zufällige Unterschiede vor Versuchsbeginn bestehen können. Dabei beachte man, daß vorher bestehende systematische Unterschiede zwischen den experimentellen Gruppen sich in einer stochastischen Abhängigkeit der unabhängigen Variablen X, welche die Gruppenzugehörigkeiten anzeigt, und einer potentiellen Störvariablen W niederschlagen würde, wie wir nun zeigen.

Ist beispielsweise

$$(5.7.1) \quad x := \begin{cases} 1, & \text{wenn die Beobachtungseinheit zur Versuchs-} \\ & \text{gruppe gehört,} \\ 0, & \text{wenn sie zur Kontrollgruppe gehört,} \end{cases}$$

so läßt sich ein Unterschied zwischen den beiden experimentellen Gruppen bezüglich des Erwartungswertes einer abhängigen Variablen Y durch die Gleichung

$$(5.7.2) \quad E(Y | X) = \alpha_{Y0} + \alpha_{YX} X, \quad \text{mit } \alpha_{YX} \neq 0$$

beschreiben. Würde bereits vor der experimentellen Behandlung ein Unterschied bezüglich der abhängigen Variablen zwischen den beiden Gruppen bestehen, so wäre nicht ohne weiteres zu entscheiden, ob der Unterschied  $E(Y | X=1) - E(Y | X=0)$ , der aus 5.7.2 folgen würde, auf die experimentelle Behandlung, oder aber auf die bereits vorher vorhandenen Unterschiede zurückzuführen ist. Bezeichnen wir die abhängige Variable vor Beginn der experimentellen Behandlung mit W, dann gälte nämlich wegen der Zweiwertigkeit von X

$$(5.7.3) \quad E(W | X) = \alpha_{W0} + \alpha_{WX} X, \quad \text{mit } \alpha_{WX} \neq 0,$$

wenn vor der experimentellen Behandlung bereits Unterschiede bezüglich des Erwartungswertes der abhängigen Variablen W bestehen.

Wenn wir von einem reglinearen Zusammenhang zwischen der Residualvariable F = Y - E(Y | X) und der abhängigen Variablen W vor der experimentellen Behandlung ausgehen, dann gälte außerdem

$$(5.7.4) \quad E(F | W) = \alpha_{F0} + \alpha_{FW} W, \quad \text{mit } \alpha_{FW} \neq 0.$$

Aus den Gleichungen 5.7.2 bis 5.7.4 folgt aber nach Theorem 5.6.2 bereits, daß Y nicht von X einfach kausal reglinear abhängig ist. Denn  $E(W | X)$  wäre nach Gleichung 5.7.3 ungleich  $E(W)$ . Eine reglineare Abhängigkeit der abhängigen Variablen W vor der experimentellen Behandlung von der unabhängigen Variablen X (siehe Gleichung 5.7.3) bei  $\alpha_{FW} \neq 0$ , würde also die Annahme einer einfachen kausalen reglinearen Abhängigkeit der Variablen Y von X falsifizieren.

Mit der experimentellen Kontrolltechnik der Randomisierung, d.h. der zufälligen Aufteilung der Beobachtungseinheiten auf die Versuchsgruppen, erreicht man, daß vor Versuchsbeginn keine systematischen Unterschiede zwischen den experimentellen Gruppen bestehen, so daß dann X und W stochastisch unabhängig sind, und die Gleichung

$$(5.7.5) \quad E(W | X) \stackrel{fs}{=} \alpha_{W0} + 0 \cdot X = E(W)$$

gilt (vgl. auch Gleichung 5.7.3), da X ja die Zugehörigkeit zu den Versuchsgruppen angezeigt. In diesem Fall ist dann die kausale Interpretierbarkeit durch die abhängige Variable vor Versuchsbeginn nicht gefährdet, wenn man voraussetzt, daß keine Interaktion im varianzanalytischen Sinn zwischen X und W besteht, die durch Randomisierung natürlich nicht ausgeschaltet werden kann.

Diesen Sachverhalt kann man etwas anders auch folgendermaßen beschreiben, wenn wir uns einen Vorgriff auf bisher noch nicht definierte Begriffe erlauben. Angenommen die direkten kausalen reglinearen Abhängigkeiten einer Variablen Y von den Variablen X und W können durch die Gleichung

$$(5.7.6) \quad E(Y | X, W) \stackrel{fs}{=} \beta_{Y0} + \beta_{YX} X + \beta_{YW} W$$

beschrieben werden, dann impliziert die stochastische Unabhängigkeit der Variablen X und W, daß die reglineare Abhängigkeit der Variablen Y von X einfach kausal ist, d.h. es gilt dann

$$(5.7.7) \quad E(Y | X) \stackrel{fs}{=} \alpha_{Y0} + \alpha_{YX} X,$$

wobei  $\alpha_{YX} = \beta_{YX}$ , denn nach den Gleichungen A.5.7 und A.5.10 erhalten wir

$$\begin{aligned} (5.7.8) \quad E[E(Y | X, W) | X] &\stackrel{fs}{=} E(Y | X) \stackrel{fs}{=} \\ &\stackrel{fs}{=} \beta_{Y0} + \beta_{YX} X + \beta_{YW} E(W | X) \stackrel{fs}{=} \\ &\stackrel{fs}{=} [\beta_{Y0} + \beta_{YW} E(W)] + \beta_{YX} X \\ &= \alpha_{Y0} + \alpha_{YX} X, \end{aligned}$$

da aus der stochastischen Unabhängigkeit von X und W folgt;  $E(W | X) \stackrel{fs}{=} E(W)$  (siehe Theorem A.5.5).

Durch Randomisierung, d.h. die zufällige Aufteilung der Beobachtungseinheiten auf die experimentellen Gruppen, kann man für alle potentielle Störvariablen W erreichen, daß eine stochastische Unabhängigkeit mit X besteht, denn die obige Argumentation ist nicht nur dann gültig, wenn W die abhängige Variable gemessen vor der experimentellen Behandlung ist, sondern sie gilt für jede beliebige potentielle Störvariable W. Durch die Kontrolltechnik der Randomisierung wird aber von vornherein sichergestellt, daß die im Theorem 5.6.1 formulierte notwendige Bedingung einfach kausaler reglinearer Abhängigkeit, nämlich Gleichung 5.6.1, für viele dieser potentiellen Störvariablen erfüllt ist.

Randomisierung ist jedoch keine Garantie dafür, daß eine einfache kausale reglineare Abhängigkeit besteht, wie auch das Beispiel ‚Drogen und Aktivierung‘ zeigt, in dem der Fall einer Interaktion im varianzanalytischen Sinn besprochen wurde. Solche Interaktionen einer unabhängigen Variablen X (im Beispiel: der Drogenvariablen) mit einer weiteren X gleich- oder vorgeordneten Variablen W (im Beispiel: der Intro/Extraversionsvariablen) sind hinreichend dafür, daß eine reglineare Abhängigkeit einer Variablen Y von X nicht einfach kausal ist. Dabei ist es irrelevant, ob diese weitere Variable W erhoben wurde oder nicht. Wichtig ist dabei nur, ob eine solche Interaktion in der betreffenden Situationsklasse besteht oder nicht.

Neben der Kontrolltechnik der Randomisierung lassen sich eventuell vorher bestehende Unterschiede zwischen den experimentellen Gruppen bezüglich des Erwartungswerts der abhängigen Variablen auch ausschalten, indem man die Gruppen bezüglich der abhängigen Variablen W vor der experimentellen Behandlung parallelisiert, d.h. durch gezielte Aufteilung der Versuchspersonen auf die experimentellen Gruppen, bezüglich der Variablen W gleichmacht, so daß dann nicht mehr  $\alpha_{wx} \neq 0$  gilt (siehe Gleichung 5.7.3). Gegenüber der Randomisierung hat die Parallelisierung den Nachteil, daß man mit dieser Kontrolltechnik nur für diejenige potentielle Störvariable W, bezüglich der parallelisiert wird, erreicht, daß die Gleichung 5.6.1 erfüllt ist. Mit der Randomisierung hingegen kann man erreichen, daß die Gleichung 5.6.1 für alle Störvariablen W erfüllt ist, wobei keine davon tatsächlich beobachtet werden muß.

## 5.8 Zusammenfassende Bemerkungen

In diesem Abschnitt wurden die formalen Eigenschaften des Begriffs der einfachen kausalen reglinearen Abhängigkeit einer stochastischen Variablen Y von einer zweiten stochastischen Variablen X behandelt. Es wurde gezeigt, daß die verschiedenen experimentellen Kontrolltechniken auf diese Eigenschaften basiert werden können, ebenso wie Strategien zur Falsifizierung von Aussagen über einfache kausale reglineare Abhängigkeit in nichtexperimentellen Situationen.

Als wichtigste formale Eigenschaft wurde bewiesen, daß die Unabhängigkeitsbedingung

$$(5.8.1) \quad E[E(F | W) | X] = 0$$

für alle potentiellen Störvariablen  $W$  gilt, wenn  $Y$  von  $X$  einfach kausal reglinear abhängig ist. Diese ist z.B. dann erfüllt, wenn  $X$  und  $W$  stochastisch unabhängig sind.

Auf diese Eigenschaft lassen sich verschiedene experimentelle Kontrolltechniken basieren. Mit der Randomisierung, d.h. der zufälligen Aufteilung der Beobachtungseinheiten auf die beiden experimentellen Gruppen<sup>19</sup>, die durch die unabhängige Variable  $X$  repräsentiert werden, legt man ein Experiment so an, daß  $X$  und alle potentielle Störvariablen  $W$  stochastisch unabhängig sind, so daß dann die Gleichung 5.8.1 erfüllt ist, die ja eine notwendige Bedingung einfacher kausaler reglinearer Abhängigkeit der Variablen  $Y$  von  $X$  ist. Mit der Parallelisierung bezüglich einer bestimmten potentiellen Störvariablen  $W$  macht man die experimentellen Gruppen bezüglich  $W$  gleich. Da  $X$  die Gruppenzugehörigkeit repräsentiert, wird damit erreicht, daß  $X$  und die betreffende Störvariable  $W$  stochastisch unabhängig sind. Das gleiche Ziel erreicht man auch mit der Konstanthaltung einer potentiellen Störvariablen  $W$ , eine Technik, die auch dazu führt, daß für diese Variable  $W$  dann die Gleichung 4.4.2 gilt. Konstanthaltung ist also die einzige dieser Techniken, die dazu führt, daß auch das in Gleichung 4.4.2 enthaltene Interaktionsverbot zwischen  $X$  und  $W$  für die konstantgehaltene Variable  $W$  erfüllt ist.

## 6. Münze und Elektromagnet mit zwei Schaltern

### 6.1 Einleitende Bemerkungen

In diesem Abschnitt soll die Diskussion der Fragen des Geltungsbereichs oder der externen Validität (Campbell & Stanley, 1963) einer Aussage über kausale Abhängigkeiten vorzubereitet werden. Dazu wird das Beispiel ‚Münzen und Elektromagnet‘ abgewandelt, indem wir einen Elektromagneten betrachten, der zwei Schalter hat. Mit jedem dieser Schalter kann man eine gewisse Feldstärke des Magneten hinzuschalten. Wenn beide Schalter auf ‚Aus‘ stehen, so ist der Elektromagnet völlig ausgeschaltet, steht einer der beiden Schalter auf ‚An‘, so hat der Magnet dieselbe Feldstärke, wie im Beispiel des Abschnitts 2, und sind beide Schalter an, so ist die Feldstärke doppelt so groß. An diesem

---

<sup>19)</sup> Bei mehr als zwei Gruppen braucht man zur Kodierung der Gruppenzugehörigkeiten mehr als nur eine unabhängige Variable  $X$ , worauf wir später ausführlicher eingehen werden.

Beispiel wollen wir überprüfen, ob die dabei auftretenden stochastischen Abhängigkeiten der Münzwurfergebnisse von den Zuständen des Elektromagneten durch ein kausales reglineares Modell beschrieben werden können. Als Alternative wird das logitlineare Modell angeführt, das aus der Literatur zur Analyse binärer Daten bekannt ist.

## 6.2 Beschreibung des Beispiels

Im Gegensatz zum Beispiel des Abschnitts 2 können wir uns hier mit dem Werfen einer einzigen Münze begnügen, die auf einer Seite aus Plastik und auf der anderen Seite aus Metall besteht. Diese Münze wird auf eine Platte geworfen, welche die Eigenschaften eines Elektromagneten hat. Dieser kann in einem von vier verschiedenen Zuständen sein: Während die Münze geworfen wird, sei der Elektromagnet mit der Wahrscheinlichkeit von 0.25 im ersten Zustand, in dem er ausgeschaltet ist. Außerdem nehmen wir an, daß unter diesem Zustand, die bedingte Wahrscheinlichkeit, daß die Münze auf die Metallseite fällt, 0.5 beträgt (siehe die erste Spalte in Tabelle 6.2.1).

Tabelle 6.2.1: Die vier Spalten repräsentieren vier Situationsklassen, die jeweils mit Wahrscheinlichkeit 0.25 auftreten, in denen beide Schalter des Elektromagneten aus, einer der Schalter an und beide Schalter an sind, während die Münze geworfen wird. Die angegebenen Zahlen sind die bedingten Wahrscheinlichkeiten, daß die Münze auf die Metall- bzw. Plastikseite fällt, gegeben eine der vier Bedingungen.

		Schalter 1			
		aus		an	
		Schalter 2		Schalter 2	
Münze fällt auf	Metallseite	aus	an	aus	an
		0.5	0.9	0.9	0.9878
		Plastikseite	0.5	0.1	0.0122

Ebenfalls jeweils mit Wahrscheinlichkeit 0.25 sei der Elektromagnet im zweiten und dritten Zustand, daß nämlich der erste Schalter an und der zweite aus bzw. der erste aus und der zweite Schalter an ist, während die Münze geworfen wird. Für beide Zustände nehmen wir an, daß die bedingte Wahrscheinlichkeit, daß die Münze auf die Metallseite fällt, 0.9 beträgt.

Im vierten Zustand schließlich, der ebenfalls mit Wahrscheinlichkeit 0.25 eintreten möge, sind beide Schalter an. Dies kann nun aber nicht bedeuten, daß der Effekt  $0.9 - 0.5 = 0.4$ , der festzustellen ist, wenn ein Schalter an ist (siehe auch Abschnitt 2), nun einfach verdoppelt sein kann, da sonst das Intervall  $[0,1]$  für die bedingte Wahrscheinlichkeit überschritten wäre, denn  $0.5 + 0.4 + 0.4 > 1.0$ . Die Erhöhung der bedingten Wahrscheinlichkeit von 0.9, unter der Bedingung, daß der erste Schalter an ist, auf eine bedingte Wahrscheinlichkeit für die Bedingung, daß beide Schalter an sind, muß also so erfolgen, daß der Betrag von 1.0 nicht überschritten wird. Wenn man bedenkt, daß sich die physikalischen Feldstärken der mit den beiden Schaltern kontrollierten Magnetfeldern aufaddieren, so ist es wohl plausibel, wenn wir die fehlende bedingte Wahrscheinlichkeit nach einem Modell festlegen, in dem sich ebenfalls die Effekte aufaddieren. Ein solches Modell ist das logitlineare Modell, nach welchem die in Tabelle 6.2.1 angegebene bedingte Wahrscheinlichkeit von 0.9878, gegeben beide Schalter sind an, berechnet wurde.

Das logitlineare Modell ist aus der Literatur zur statistischen Analyse binärer Daten (siehe z.B. Bishop, Fienberg & Holland, 1975, Coleman, 1981, Cox, 1970 oder Goodman, 1978), aber auch in der Literatur zur Konstruktion psychometrischer Tests (siehe z.B. Fischer, 1974, Lord & Novick, 1968, Lumsden, 1976 und Rasch, 1960) wohlbekannt. Definieren wir die Münzvariable

$$(6.2.1) \quad Y_1 := \begin{cases} 1, & \text{wenn die Münze auf die Metallseite,} \\ 0, & \text{wenn sie auf die Plastikseite fällt,} \end{cases}$$

die erste Magnetvariable

$$(6.2.2) \quad X_1 := \begin{cases} 1, & \text{wenn der erste Schalter an,} \\ 0, & \text{wenn er aus ist,} \end{cases}$$

die zweite Magnetvariable

$$(6.2.3) \quad X_2 := \begin{cases} 1, & \text{wenn der zweite Schalter an,} \\ 0, & \text{wenn er aus ist,} \end{cases}$$

und das Ereignis

$$(6.2.4) \quad B := \{\omega \in \Omega: Y_1(\omega) = 1\},$$

daß die Münze auf die Metallseite fällt, so können die bedingten Wahrscheinlichkeiten in Tabelle 6.2.1 durch das folgende logitlineare Modell erzeugt werden:

$$(6.2.5) \quad E(Y_1 | X_1, X_2) \stackrel{f_s}{=} P(B | X_1, X_2) \stackrel{f_s}{=} \frac{\exp(\beta_{10} + \beta_{11} X_1 + \beta_{12} X_2)}{1 + \exp(\beta_{10} + \beta_{11} X_1 + \beta_{12} X_2)} =$$

$$= \frac{\exp(2.1972 \cdot X_1 + 2.1972 \cdot X_2)}{1 + \exp(2.1972 \cdot X_1 + 2.1972 \cdot X_2)}.$$

Wie man sich leicht überzeugen kann, kann man mit dieser Gleichung die Daten der Tabelle 6.2.1 bis auf Rundungsfehler reproduzieren. So resultiert aus Gleichung 6.2.5 z.B. die bedingte Wahrscheinlichkeit

$$(6.2.6) \quad P(B | X_1=1, X_2=1) = \frac{\exp(4.3944)}{1 + \exp(4.3944)} = 0.9878,$$

und auf entsprechende Weise erhält man die anderen bedingten Wahrscheinlichkeiten in Tabelle 6.2.1, die auch so gewählt wurden, daß sie mit der Gleichung 6.2.5 beschrieben werden können.<sup>20)</sup>

### 6.3 Reglineare Abhängigkeit

Wir versetzen uns nun in die Lage eines Forschers, der die in Tabelle 6.2.1 angegebenen Wahrscheinlichkeiten als relative Häufigkeiten einer langen Versuchsreihe gefunden hat, in welcher der Versuch, die Münze zu werfen, während der Elektromagnet mit den angegebenen Wahrscheinlichkeiten in einem der vier Zustände ist, sehr oft durchgeführt wurde.

Als erstes wollen wir dabei untersuchen, ob sich die stochastische Abhängigkeit der Münzwurfvariablen  $Y_1$  von der ersten Magnetvariablen  $X_1$  als einfache kausale reglineare Abhängigkeit beschreiben lässt. Dabei legen wir den Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  grunde, der alle vier in Tabelle 6.2.1 aufgeführten Situationsklassen, die durch die vier Zustände des Elektromagneten charakterisiert sind, als Unterklassen enthält, d.h.  $(\Omega, \mathcal{A}, P)$  soll den gesamten Versuch repräsentieren.

Die bedingte Wahrscheinlichkeit, daß die Münze auf die Metallseite fällt, gegeben der erste Schalter ist an, beträgt dann 0.9439, und die entsprechende Wahrscheinlichkeit, gegeben der erste Schalter ist aus, beträgt dann 0.7. Diese Daten sind noch einmal in einer etwas anderen Art als in Tabelle 6.2.1 in der Tabelle 6.3.1 dargestellt, aus der nun auch zu ersehen ist, daß die beiden Magnetvariablen  $X_1$  und  $X_2$ , welche die Schalterstellungen anzeigen, stochastisch unabhängig sind.

<sup>20)</sup> Ob sich die Daten eines entsprechenden echten physikalischen Experiments tatsächlich durch ein logitlineares Modell beschreiben lassen, ist wohl eine empirische Frage, und hier letztlich irrelevant, da dieses Beispiel nur dazu dient, Grundgedanken kausaler reglinearer und logitlinearer Modelle zu erläutern, und die Diskussion der Fragen externer Validität im nächsten Abschnitt vorzubereiten.

Tabelle 6.3.1: Vier Situationsklassen, die jeweils mit Wahrscheinlichkeit 0.25 auftreten, in denen beide Schalter des Elektromagneten aus, einer der Schalter an und beide Schalter an sind, während die Münze geworfen wird. Die angegebenen Zahlen im Innern der Tabelle sind die bedingten Wahrscheinlichkeiten, daß die Münze auf die Metallseite fällt gegeben eine der vier Bedingungen. In Klammern sind die Wahrscheinlichkeiten für das Auftreten der vier Situationsklassen und deren Vereinigungen angegeben.

		Schalter 2		
		an	aus	
Schalter 1	an	0.9878 (0.25)	0.9 (0.25)	0.9439 (0.5)
	aus	0.9 (0.25)	0.5 (0.25)	0.7 (0.5)
		0.9439 (0.5)	0.7 (0.5)	(1.0)

Wegen der Zweiwertigkeit von  $X_1$  läßt sich die stochastische Abhängigkeit der Münzvariablen  $Y_1$  von  $X_1$  durch

$$(6.3.1) \quad E(Y_1 | X_1) = P(B | X_1) = \alpha_{10} + \alpha_{11} X_1 = 0.7 + 0.2439 \cdot X_1$$

als reglineare Abhängigkeit beschreiben, wobei wir die numerischen Werte der Koeffizienten aus den beiden Gleichungen

$$(6.3.2) \quad P(B | X_1=1) = \alpha_{10} + \alpha_{11} = 0.9439$$

und

$$(6.3.3) \quad P(B | X_1=0) = \alpha_{10} = 0.7$$

erhalten, die beide aus Gleichung 6.3.1 nach den Gleichungen A.5.2, A.4.6 und A.4.3 des Anhangs folgen.

In diesem Beispiel können wir eine eindeutige zeitliche Geordnetheit der Variablen voraussetzen, denn die beiden Magnetvariablen sind der Versuchsanlage gemäß der Münzvariablen vorgeordnet. Falls in diesem Beispiel die Invarianzbedingung für die durch Gleichung 6.3.1 beschriebenen reglinearen Abhängigkeit gegeben wäre, könnten wir diese Abhängigkeit auch kausal interpretieren und den Koeffizienten  $\alpha_{11}$  als einfachen kausalen reglinearen Effekt von  $X_1$  auf  $Y_1$  bezeichnen. Die Bedingung der Invarianz ist jedoch nicht erfüllt, wie wir nun anhand des Theorems 5.4.1 verifizieren.

Offensichtlich besteht in diesem Beispiel eine Interaktion im varianzanalytischen Sinn zwischen der ersten und der zweiten Magnetvariablen  $X_1$  bzw.  $X_2$ . Ein multiples reglineares Modell mit Interaktion, welches die Abhängigkeiten

von  $Y_1$  unter allen vier Zuständen des Elektromagneten beschreibt, ist demnach

$$(6.3.4) \quad E(Y_1 | X_1, X_2) = P(B | X_1, X_2) = \\ = 0.5 + 0.4 \cdot X_1 + 0.4 \cdot X_2 + (-0.3122) \cdot X_1 \cdot X_2,$$

wobei wir die numerischen Werte auf analoge Weise wie beim Modell mit einer einzigen unabhängigen Variablen erhalten, nur daß wir hier vier Gleichungen mit vier Unbekannten zu lösen haben.

Nicht nur mit dem logitlinearen Modell (siehe Gleichung 6.2.5), sondern auch mit diesem reglinearen Modell mit Interaktion lassen sich die stochastischen Abhängigkeiten der Münzvariablen  $Y_1$  von den beiden Magnetvariablen  $X_1$  und  $X_2$  perfekt beschreiben, denn auch mit Gleichung 6.3.4 können die in Tabelle 6.2.1 angegebenen bedingten Wahrscheinlichkeiten exakt reproduziert werden. Die in diesem reglinearen Modell vorkommenden Parameter lassen sich jedoch nicht kausal interpretieren, da die in Definition 4.4.1 formulierte Invarianzbedingung nicht erfüllt ist. Zum einen bleibt nämlich der Parameter  $\alpha_{11} = 0.2439$  aus Gleichung 6.3.1 bei der Erweiterung von  $E(Y_1 | X_1)$  auf  $E(Y_1 | X_1, X_2)$  nicht invariant, denn in Gleichung 6.3.4 hat  $X_1$  nicht den Koeffizienten 0.2439, sondern 0.4, und zum zweiten kommt in Gleichung 6.3.4 der Interaktionsparameter -0.3122 vor, der nach Theorem 5.4.1 gleich Null sein müßte. Schon eine dieser beiden Tatsachen aber genügt, um die Annahme zu falsifizieren, daß  $Y_1$  von  $X_1$  auf dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  einfach kausal reglinear abhängig ist, wobei  $(\Omega, \mathcal{A}, P)$  den gesamten Versuch repräsentiert.

Der in Gleichung 6.3.4 vorkommende Interaktionsparameter ist in gewisser Weise artifiziell, da es ein anderes Modell gibt, welches ohne einen solchen Parameter auskommt, nämlich das logitlineare Modell der Gleichung 6.2.5. In diesem Modell verhalten sich die beiden Parameter additiv, was in besserer Übereinstimmung mit unserem physikalischen Wissen über die Additivität der Feldstärken von Magneten steht.

Dieses Beispiel illustriert noch einmal die in der Definition 4.4.1 formulierte Eigenschaft, daß eine reglineare Abhängigkeit einer Variablen  $Y$  von  $X$  immer dann nicht einfach kausal ist, wenn eine Interaktion von  $X$  mit potentiellen Störvariablen  $W$  (im Beispiel: der zweiten Magnetvariablen) besteht. Ein möglicher Weg wäre nun, die Aussage, daß die Münzvariable  $Y_1$  einfach kausal reglinear von der ersten Magnetvariablen  $X_1$  abhängig ist, auf jeweils eine der beiden Bedingungen, daß der zweite Schalter an bzw. aus ist, einzuschränken, so daß für jede dieser beiden Situationen ein unterschiedlicher einfacher kausaler reglinearer Effekt der ersten Magnetvariablen auf die Münzvariable angenommen wird, worauf wir ausführlicher im Abschnitt 7 eingehen werden. Die Bedingung der Invarianz könnte dann jeweils innerhalb dieser beiden Situationsklassen überprüft werden, indem man nach anderen potentiellen Störvariablen sucht und prüft, ob für diese die Invarianzbedingung erfüllt ist.

## 6.4 Logitlineare Abhangigkeit

Angenommen, wir wubten nichts von der zweiten Magnetvariablen  $X_2$  und wollten die Abhangigkeit der Munzvariablen  $Y_1$  von der ersten Magnetvariablen  $X_1$  durch ein einfaches logitlineares Modell beschreiben. Bezeichnen wir das Ereignis, da die Munze auf die Metallseite fallt, weiterhin mit

$$(6.4.1) \quad B := \{\omega \in \Omega : Y_1(\omega) = 1\},$$

so gilt dann die Gleichung

$$(6.4.2) \quad E(Y_1 | X_1) \stackrel{\text{fs}}{=} P(B | X_1) \stackrel{\text{fs}}{=} \frac{\exp(\gamma_{10} + \gamma_{11} X_1)}{1 + \exp(\gamma_{10} + \gamma_{11} X_1)}.$$

Aus den daraus folgenden beiden bedingten Wahrscheinlichkeiten  $P(B | X_1 = 0)$  und  $P(B | X_1 = 1)$  erhalten wir nach einigen algebraischen Umformungen die folgenden beiden Gleichungen:

$$(6.4.3) \quad \gamma_{10} = \ln \left[ \frac{P(B | X_1=0)}{1 - P(B | X_1=0)} \right] = \ln [0.7/(1 - 0.7)] = 0.8473$$

und

$$(6.4.4) \quad \begin{aligned} \gamma_{10} + \gamma_{11} &= \ln \left[ \frac{P(B | X_1=1)}{1 - P(B | X_1=1)} \right] = \\ &= \ln [0.9439/(1 - 0.9439)] = 2.8229. \end{aligned}$$

Daraus ergibt sich der Koeffizient  $\gamma_{11} = 1.9756$ , der nicht mit  $\beta_{11} = 2.1972$  aus Gleichung 6.2.5 identisch ist.

Hierin besteht ein wichtiger *Unterschied* zum reglinearen *Modell*. Obwohl  $X_1$  und  $X_2$  stochastisch unabhangig sind (siehe Tabelle 6.3.1), darf man  $X_2$  im logitlinearen Fall offenbar nicht ignorieren, denn sonst kann man den tatsachlichen kausalen Effekt (hier:  $\beta_{11} = 2.1972$ ) nicht korrekt bestimmen. Im Fall einer einfachen kausalen reglinearen Abhangigkeit konnte man  $X_2$  ignorieren, und wurde dennoch den einfachen kausalen reglinearen Effekt von  $X_1$  auf  $Y_1$  korrekt ermitteln, falls  $X_1$  und  $X_2$  stochastisch unabhangig sind. Diese besonders vorteilhafte Eigenschaft, auf die wir die Kontrolltechnik der Randomisierung basieren konnen (siehe Abschnitt 5.7), gilt im Fall logitlinearer Abhangigkeit offenbar nicht. Bei multipler logitlinearer Abhangigkeit (siehe Gleichung 6.2.5) ist Randomisierung anscheinend kein Verfahren, das dazu beitragen kann, eine kausale Interpretation einer solchen logitlinearen Abhangigkeit zu ermoglichen. Offenbar mu man im logitlinearen Fall statt dessen auf die Kontrolltechnik der Konstanthalzung (hier: der zweiten Magnetvariablen) zuruckgreifen, worauf wir ausfuhrlicher im Abschnitt 7 eingehen.

## 6.5 Zusammenfassende Bemerkungen

In diesem Abschnitt wurde ein Beispiel dargestellt, in dem das Ergebnis des Werfens einer Münze durch zwei Schalter eines Elektromagneten beeinflußt werden kann. In diesem Beispiel läßt sich die Abhängigkeit der Münzvariablen  $Y_1$  von den beiden Magnetvariablen  $X_1$  und  $X_2$ , welche jeweils die Schalterstellungen angeben, sowohl durch ein reglineares Modell mit Interaktion zwischen den beiden Magnetvariablen beschreiben, als auch durch ein logitlineares Modell, in dem keine solche Interaktion vorkommt. Die Interaktion im reglinearen Modell kann daher in diesem Beispiel als modellbedingt oder artifiziell bezeichnet werden. Wegen der Interaktion zwischen  $X_1$  und  $X_2$  ist in diesem Beispiel die Münzvariable  $Y_1$  nicht von einer der Magnetvariablen  $X_1$  oder  $X_2$  einfach kausal reglinear abhängig.

Versucht man die Abhängigkeit der Münzvariablen  $Y_1$  z.B. von der ersten Magnetvariablen  $X_1$  durch ein einfaches logitlineares Modell zu beschreiben, so erhält man dabei einen anderen Parameter als denjenigen, mit dem im Modell mit beiden Magnetvariablen  $X_1$  und  $X_2$  die Daten erzeugt wurden, obwohl die beiden Magnetvariablen stochastisch unabhängig sind. Hierin besteht ein wichtiger Unterschied zwischen logitlinearer und reglinearer Abhängigkeit, denn beim reglinearen Fall ist die stochastische Unabhängigkeit von  $X_1$  und  $X_2$  hinreichend dafür, daß  $X_2$  die Beziehung zwischen  $Y$  und  $X_1$  nicht modifiziert, falls auch keine Interaktion zwischen  $X_1$  und  $X_2$  besteht (siehe Abschnitt 5.6).

In diesem Beispiel kann weder die einfache reglineare, noch die einfache logitlineare Abhängigkeit als kausal bezeichnet werden, wenn man einfache kausale logitlineare Abhängigkeit analog zur einfachen kausalen reglinearen Abhängigkeit definieren würde. Als Ausweg bleibt in einem solchen Fall nur die Einschränkung des Geltungsbereichs der Modelle mit einer einzigen unabhängigen Variablen auf jeweils eine der beiden Situationen, daß der zweite Schalter des Elektromagneten an oder aber aus ist, worauf wir ausführlicher u. a. im folgenden Abschnitt eingehen werden.

## 7. Externe Validität

### 7.1 Einleitende Bemerkungen

Der Begriff der externen Validität wurde neben dem der internen Validität (siehe Abschnitt 3) als das zweite fundamentale Gütekriterium für eine empirische Untersuchung von Campbell (siehe z.B. Campbell, 1957, Campbell & Stanley, 1963) eingeführt: „Externe Validität fragt nach der Generalisierbar-

keit: Auf welche Populationen, situativen Bedingungen, Behandlungs- und Meßvariablen kann dieser Effekt verallgemeinert werden?“ (Campbell & Stanley, 1963, S. 175, Übersetzung durch den Autor).

Sieht man von den bereits von Gadenne (1976) und später auch von Bredenkamp (1979, 1980) kritisierten induktivistischen Formulierungen ab, so verbergen sich in der Tat hinter dieser Bezeichnung zwei Fragen, die einer expliziten Diskussion bedürfen: Zum einen die Frage nach dem Geltungsbereich einer Aussage über eine kausale stochastische Abhängigkeit, und zum anderen die Frage nach der Variablenvalidität, also die Frage: „Wie gut repräsentieren die Variablen einer Untersuchung das jeweils Gemeinte?“ (Bredenkamp, 1980, S. 31).

Zur Lösung der zweiten Frage ist die Theorie latenter Variablen von Bedeutung, über die Moosbrugger (1982) in diesem Band einen Überblick gibt. In diesem Abschnitt werden wir nur auf die Frage nach dem Geltungsbereich näher eingehen. Bezüglich des Problems der Variablenvalidität müssen wir dabei voraussetzen, daß die jeweils betrachteten Variablen tatsächlich die in einer Untersuchung gemeinten sind. Dabei müssen diese Variablen keineswegs direkt beobachtbar oder ‚manifest‘ sein, statt dessen kann es sich ebensogut um ‚latente‘, nicht direkt beobachtbare Variablen handeln.

Auch die verbliebene Frage nach dem Geltungsbereich kann in zwei verschiedene Aspekte untergliedert werden, den der Situationsvalidität und den der Populationsvalidität, die wir in den folgenden beiden Abschnitten behandeln werden. Im letzten Abschnitt zeigen wir dann, wie zwei verschiedene Modelle hinsichtlich ihrer ‚externen Validität‘ bzw. ihres Geltungsbereichs miteinander verglichen werden können.

## 7.2 Situationsvalidität

Der erste Aspekt der externen Validität, den wir hier behandeln wollen, ist der der „Situationsvalidität“ (Hager & Westermann, 1982) oder der „ökologischen Validität“ (Bredenkamp, 1979, 1980; Brunswik, 1956). Bei diesem Gesichtspunkt der externen Validität handelt es sich um die Frage, für welche Situationsklassen oder „situativen Bedingungen“ eine kausale stochastische Abhängigkeit behauptet werden kann. Wir zeigen nun, wie diese Fragen in der hier vorgestellten Theorie behandelt werden können.

Dabei gehen wir von dem im Abschnitt 6 beschriebenen Beispiel aus. Dort hatten wir festgestellt, daß die reglineare Abhängigkeit der Münzvariablen  $Y_1$  von der ersten Magnetvariablen  $X_1$  nicht einfach kausal ist, wenn wir den Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  zugrunde legen, der die Situationsklasse repräsentiert, in welcher der zweite Schalter mit Wahrscheinlichkeit 0.5 an

bzw. aus ist, und als einen Ausweg haben wir dort erwogen, den Geltungsbereich der Aussage jeweils auf eines der beiden Ereignisse

$$(7.2.1) \quad A := \{\omega \in \Omega: X_2(\omega) = 1\},$$

daß der zweite Schalter an ist, bzw.

$$(7.2.2) \quad \bar{A} := \{\omega \in \Omega: X_2(\omega) = 0\},$$

daß der zweite Schalter aus ist, einzuschränken, oder mit anderen Worten, den zweiten Schalter konstant zu halten. Formal können wir dies dadurch ausdrücken, daß wir nicht mehr den Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$ , sondern die Wahrscheinlichkeitsräume  $(\Omega, \mathcal{A}, P_A)$  bzw.  $(\Omega, \mathcal{A}, P_{\bar{A}})$  zugrunde legen, wobei die Wahrscheinlichkeitsmaße  $P_A$  und  $P_{\bar{A}}$  durch

$$(7.2.3) \quad P_A(B) := P(B | A) := P(A \cap B) / P(A),$$

für alle  $B \in \mathcal{A}$ ,  $P(A) > 0$ ,

bzw.

$$(7.2.4) \quad P_{\bar{A}}(B) := P(B | \bar{A}) = P(\bar{A} \cap B) / P(\bar{A}),$$

für alle  $B \in \mathcal{A}$ ,  $P(\bar{A}) > 0$ ,

definiert sind, vorausgesetzt, daß die Wahrscheinlichkeiten von  $A$  und  $\bar{A}$  größer Null sind.

Bezeichnen wir mit

$$(7.2.5) \quad B := \{\omega \in \Omega: Y_1(\omega) = 1\}$$

das Ereignis, daß die Münze auf die Metallseite fällt, dann gilt beispielsweise unter der Bedingung, daß der zweite Schalter aus ist, für die bedingte Erwartung der Münzvariablen  $Y_1$  unter der ersten Magnetvariablen  $X_1$  bezüglich des Maßes  $P_{\bar{A}}$  die Gleichung<sup>21)</sup>

$$(7.2.6) \quad E_{\bar{A}}(Y_1 | X_1) \underset{P_{\bar{A}}-\text{fs}}{=} P_{\bar{A}}(B | X_1) \underset{P_{\bar{A}}-\text{fs}}{=} \beta_{10} + \beta_{11} X_1 = 0.5 + 0.4 \cdot X_1,$$

wobei man die numerischen Werte der Koeffizienten aus der Tabelle 6.2.1 durch Bildung der bedingten Erwartungswerte  $E_{\bar{A}}(Y_1 | X_1=0) = P_{\bar{A}}(B | X_1=0)$  und  $E_{\bar{A}}(Y_1 | X_1=1) = P_{\bar{A}}(B | X_1=1)$  und Lösung der resultierenden beiden Gleichungen

$$(7.2.7) \quad \beta_{10} = P_{\bar{A}}(B | X_1=0) = P(B | X_1=0, X_2=0) = 0.5$$

<sup>21)</sup> Das Zeichen,  $\underset{P_{\bar{A}}-\text{fs}}{=}$  in Gleichung 7.2.6 heißt „fast sicher bezüglich des Wahrscheinlichkeitsmaßes  $P_{\bar{A}}$ “ (siehe die Bemerkungen zu Definition A.5.1).

und

$$(7.2.8) \quad \beta_{10} + \beta_{11} = P_{\bar{A}}(B \mid X_1=1) = P(B \mid X_1=1, X_2=0) = 0.9$$

nach den zwei unbekannten Koeffizienten erhält.

Auf analoge Weise resultiert dann für die Bedingung A die folgende Gleichung für die bedingte Erwartung  $E_A(Y_1 \mid X_1)$  bezüglich des Wahrscheinlichkeitsmaßes  $P_A$

$$(7.2.9) \quad E_A(Y_1 \mid X_1)_{P_A=\text{fs}} = 0.9 + 0.0878 \cdot X_1.$$

Dies bestätigt noch einmal das in Abschnitt 6.3 gefundene Ergebnis, daß  $Y_1$  von  $X_1$  nicht einfach kausal reglinear abhängig ist, denn die Koeffizienten aus den Gleichungen 7.2.6 und 7.2.9 sind nicht identisch, wie es der Definition der einfachen kausalen reglinearen Abhängigkeit gemäß der Fall sein müßte. Der Koeffizient  $\alpha_{11} = 0.2439$  aus Gleichung 6.3.1 lässt sich also nicht als einfacher kausaler reglinearer Effekt von  $X_1$  auf  $Y_1$  interpretieren, wenn wir den Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  zugrunde legen, der den gesamten Versuch repräsentiert; er kann lediglich als *Durchschnitt der beiden kausalen einfachen reglinearen Effekte* 0.4 und 0.0878 aus den Gleichungen 7.2.6 und 7.2.9 bezeichnet werden.

Schränken wir den Geltungsbereich unserer Aussage über eine einfache kausale reglineare Abhängigkeit der Variablen  $Y_1$  von  $X_1$  auf jeweils eine der beiden Bedingungen A bzw.  $\bar{A}$  ein, indem wir anstelle des Wahrscheinlichkeitsraums  $(\Omega, \mathcal{A}, P)$  jeweils die Wahrscheinlichkeitsräume  $(\Omega, \mathcal{A}, P_A)$  bzw.  $(\Omega, \mathcal{A}, P_{\bar{A}})$  zugrunde legen, so spräche wohl nichts dagegen, die durch die Gleichungen 7.2.6 und 7.2.9 beschriebenen Abhängigkeiten jeweils kausal zu interpretieren. Entsprechend ist aber der *Geltungsbereich dieser Aussagen auf die Bedingungen  $\bar{A}$  bzw. A eingeschränkt*, oder, etwas formaler formuliert, auf die Wahrscheinlichkeitsräume  $(\Omega, \mathcal{A}, P_{\bar{A}})$  bzw.  $(\Omega, \mathcal{A}, P_A)$ , die beide in dem größeren Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  enthalten sind (siehe die Gleichungen 7.2.3 und 7.2.4), der ja den gesamten Versuch repräsentiert.

*Die Frage, unter welchen situativen Bedingungen oder in welchen Situationsklassen eine kausale stochastische Abhängigkeit gilt, ist also gleichbedeutend damit, welcher Wahrscheinlichkeitsraum zugrundegelegt wird.* Dabei beachte man, daß innerhalb der hier vorgestellten Theorie die Frage nach der Situationsvalidität nur ein Aspekt der Frage nach der Kausalität einer Abhängigkeit zwischen zwei Variablen ist, der eigentlich gar nicht von der Frage nach der Kausalität zu trennen ist. Eine vollständige Aussage über eine einfache kausale reglineare Abhängigkeit lautet nämlich: „Die stochastische Variable Y ist von der stochastischen Variablen X auf dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  einfach kausal reglinear abhängig mit dem einfachen kausalen reglinearen Effekt  $\alpha_{YX}$ .“ Das Entscheidende dabei ist, daß der Geltungsbereich, den der Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  repräsentiert, selbst bereits unverzichtbarer Bestandteil der Aussage ist. Ohne den Bezug auf einen ganz bestimmten Wahrscheinlichkeitsraum ist eine solche Aussage völlig inhaltsleer, da dann die

dabei benutzten Begriffe undefiniert wären. Eine stochastische Variable beispielsweise ist ja nur für einen ganz bestimmten Wahrscheinlichkeitsraum definiert (siehe z.B. Bauer, 1974, S. 137), der in Anwendungen jeweils eine ganz bestimmte Situationsklasse repräsentiert.

Dies soll aber nicht bedeuten, daß man sich nicht fragen kann, ob die für einen Wahrscheinlichkeitsraum (eine Situationsklasse) gültige stochastische Abhängigkeit, oder genauer, der dabei gültige Steigungskoeffizient der einfachen linearen Regression, nicht auch in einem anderen Wahrscheinlichkeitsraum gültig ist. Bei der Frage nach der Situationsvalidität geht es also darum, den oder die Wahrscheinlichkeitsräume zu spezifizieren, die man bei einer Aussage über eine einfache kausale reglineare Abhängigkeit zugrunde legen will, wenn wir von einem einfachen reglinearen Modell ausgehen.

### 7.3 Populationsvalidität

Ein in formaler Hinsicht mit der ‚Situationsvalidität‘ völlig identisches Problem, ist der zweite Aspekt der externen Validität, nämlich das Problem der „Populationsvalidität“ (Bredenkamp, 1979, 1980, Hager & Westermann, 1982). Dies soll nun etwas ausführlicher erläutert werden.

Dazu modifizieren wir das im letzten Kapitel behandelte Beispiel wie folgt: Wir betrachten zwei Münzen I und II, die beide auf jeweils einer Seite aus Plastik bzw. Metall bestehen. Mit diesen Münzen wird der folgende Versuch durchgeführt: Zuerst wird aus einer Urne nach Zufall eine der beiden Münzen gezogen, so daß jede die Wahrscheinlichkeit 0.5 hat, gezogen zu werden. Dann wird die gezogene Münze auf eine Platte geworfen, welche die Eigenschaften eines ein- und ausschaltbaren Elektromagneten hat, der ebenfalls mit Wahrscheinlichkeit 0.5 an oder aus ist, und zwar unabhängig davon, welche Münze gezogen und geworfen wird (siehe Tabelle 7.3.1).

Im Gegensatz zum Beispiel des Abschnitts 2 soll es sich jedoch nicht um zwei physikalisch gleiche Münzen handeln, sondern die Münzen sollen auf der Metallseite verschieden schwer sein, mit dem Effekt, daß sie mit unterschiedlichen Wahrscheinlichkeiten auf die Metallseite fallen. Wenn der Elektromagnet aus ist, sei für die erste Münze die bedingte Wahrscheinlichkeit 0.5, auf die Metallseite zu fallen, und für die zweite betrage diese bedingte Wahrscheinlichkeit 0.9. Der Elektromagnet sei ebenso stark wie im Beispiel des Abschnitts 2, so daß die bedingte Wahrscheinlichkeit, für die Münze I auf die Metallseite zu fallen, wenn der Elektromagnet an ist, wieder 0.9 beträgt, und für die zweite Münze 0.9878. Damit haben wir die gleichen Daten wie in Tabelle 6.2.1 vorliegen, die hier jedoch inhaltlich anders strukturiert sind. Die Stelle verschiedener Situationen, die durch die Stellung des zweiten Schalters gekennzeichnet sind, wird nun von unterschiedlichen Münzen eingenommen, und statt zu fragen, in welchen Situationen eine bestimmte kausale stochastische Abhängigkeit gilt, können wir nun fragen, für welche *Population* oder

Tabelle 7.3.1: Die Spalten repräsentieren vier Situationsklassen, die jeweils mit Wahrscheinlichkeit 0.25 auftreten. In der ersten und zweiten wird Münze I gezogen und der Schalter des Elektromagneten ist aus bzw. an. In der dritten und vierten Situationsklasse wird Münze II gezogen und der Schalter ist aus bzw. an. Die angegebenen Zahlen sind die bedingten Wahrscheinlichkeiten, daß die geworfene Münze auf die Metall- bzw. Plastikseite fällt.

		Gezogen wird			
		Münze I		Münze II	
		Schalter		Schalter	
Münze fällt auf	Metallseite	aus	an	aus	an
	Plastikseite	0.5	0.9	0.9	0.9878
		0.5	0.1	0.1	0.0122

Menge von Münzen eine solche stochastische Abhängigkeit behauptet werden kann.

Es ist offensichtlich, daß wir hier auf eine Analogie zwischen Münzen und Versuchspersonen abzielen. Wie wir im Abschnitt 7.2 gesehen haben, können sich auch Münzen in unterschiedlichen Situationen verschieden ‚verhalten‘ und in diesem Abschnitt ist deutlich geworden, daß Münzen auch unterschiedliche Eigenschaften haben können, was natürlich um so mehr für Personen oder Gruppen gilt. Dies unterstreicht die Wichtigkeit, den *Wahrscheinlichkeitsraum* bei einer Aussage über eine stochastische Abhängigkeit mitanzugeben, denn dieser *repräsentiert* sowohl die Situationsklasse, für welche die Abhängigkeit behauptet wird, als auch die *Population*, wie wir im folgenden zeigen wollen. Die in der Experimentellen Psychologie übliche genaue Beschreibung der Untersuchungssituation und die genaue Charakterisierung der beteiligten Versuchspersonen, kommt dieser Forderung nach einer möglichst exakten Beschreibung des zugrunde gelegten Wahrscheinlichkeitsraums bereits weitgehend entgegen.

Wir zeigen nun, daß auch die Frage, auf welche Population (in unserem Beispiel: von Münzen) sich eine Aussage über eine kausal reglineare Abhängigkeit bezieht, durch die Angabe des zugrunde gelegten Wahrscheinlichkeitsraums beantwortet wird.

Definieren wir die Münzwurfvariable

$$(7.3.1) \quad Y_1 := \begin{cases} 1, & \text{wenn die gewählte Münze auf die Metallseite,} \\ & \\ 0, & \text{wenn sie auf die Plastikseite fällt,} \end{cases}$$

die Magnetvariable

$$(7.3.2) \quad X_1 := \begin{cases} 1, & \text{wenn der Elektromagnet an-,} \\ 0, & \text{wenn er ausgeschaltet ist,} \end{cases}$$

die Münzwahlvariable

$$(7.3.3) \quad X_1 := \begin{cases} 1, & \text{wenn Münze I,} \\ 0, & \text{wenn Münze II gewählt wird,} \end{cases}$$

und das Ereignis

$$(7.3.4) \quad B := \{\omega \in \Omega : Y_1(\omega) = 1\},$$

daß die geworfene Münze auf die Metallseite fällt, so gelten für diese Variablen, wenn auch mit anderer Interpretation, dieselben Gleichungen wie in den Abschnitten 6.2 bis 6.4.

Wir können uns auch hier die Frage stellen, ob die reglineare Abhängigkeit der Münzwurfvariablen  $Y_1$  von der Magnetvariablen  $X_1$  einfach kausal ist, wenn wir den Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  zugrunde legen, der den gesamten Versuch repräsentiert, der also diejenige Situationsklasse beschreibt, in der die beiden Münzen jeweils die Wahrscheinlichkeit 0.5 haben, aus der Urne gezogen zu werden. Auch für dieses Beispiel finden wir, daß die in Definition 4.4.1 formulierte Invarianzbedingung nicht erfüllt ist, denn es gilt auch hier die Gleichung

$$(7.3.5) \quad E(Y_1 | X_1, X_2) \underset{fs}{=} P(B | X_1, X_2) \underset{fs}{=} \\ \underset{fs}{=} 0.5 + 0.4 \cdot X_1 + 0.4 \cdot X_2 + (-0.3122) \cdot X_1 \cdot X_2.$$

Da der Interaktionsparameter -0.3122 und nicht 0 beträgt, ist auch hier zu schließen, daß  $Y_1$  von  $X_1$  nicht einfach kausal reglinear abhängig ist.

Wir können nun den Geltungsbereich der Aussage jeweils auf eine der beiden Bedingungen

$$(7.3.6) \quad \bar{A} := \text{Münze I wird gewählt}$$

bzw.

$$(7.3.7) \quad A := \text{Münze II wird gewählt}$$

einschränken, oder mit anderen Worten, auf jeweils eine der beiden Münzen begrenzen. Formal drücken wir dies dadurch aus, daß wir nicht mehr den Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$ , sondern die Wahrscheinlichkeitsräume  $(\Omega, \mathcal{A}, P_A)$  bzw.  $(\Omega, \mathcal{A}, P_{\bar{A}})$  zugrunde legen, wobei die Wahrscheinlichkeitsmaße  $P_A$  und  $P_{\bar{A}}$  analog wie im letzten Abschnitt definiert sind. In dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P_{\bar{A}})$ , d.h. unter der Bedingung, daß Münze I ge-

wählt wird, gilt dann für die bedingte Erwartung der Münzwurfvariablen  $Y_1$  unter der Magnetvariablen  $X_1$  die Gleichung

$$(7.3.8) \quad E_{\bar{A}}(Y_1 | X_1) \underset{P_{\bar{A}}-\text{fs}}{\equiv} P_{\bar{A}}(B | X_1) \underset{P_{\bar{A}}-\text{fs}}{\equiv} \beta_{10} + \beta_{11} X_1 = 0.5 + 0.4 \cdot X_1,$$

wobei man die numerischen Werte der Koeffizienten durch Bildung der bedingten Erwartungswerte  $E_{\bar{A}}(Y_1 | X_1=0) = P_{\bar{A}}(B | X_1=0)$  und  $E_{\bar{A}}(Y_1 | X_1=1) = P_{\bar{A}}(B | X_1=1)$  und Lösung der resultierenden beiden Gleichungen nach den zwei unbekannten Koeffizienten erhält (siehe Abschnitt 7.2).

Auf analoge Weise resultiert dann für die Bedingung A die bedingte Erwartung  $E_A(Y_1 | X_1)$  bezüglich des Wahrscheinlichkeitsmaßes  $P_A$

$$(7.3.9) \quad E_A(Y_1 | X_1) \underset{P_A-\text{fs}}{\equiv} 0.9 + 0.0878 \cdot X_1.$$

Damit haben wir noch einmal bestätigt, daß  $Y_1$  von  $X_1$  nicht einfach kausal reglinear abhängig ist, denn die Koeffizienten von  $X_1$  aus den Gleichungen 7.3.8 und 7.3.9 sind nicht identisch, wie es der Definition der einfachen kausalen reglinearen Abhängigkeit gemäß der Fall sein müßte. Auch hier läßt sich also der Koeffizient  $\alpha_{11} = 0.2439$  aus Gleichung

$$(7.3.10) \quad E(Y_1 | X_1) \underset{\text{fs}}{\equiv} \alpha_{10} + \alpha_{11} X_1 = 0.7 + 0.2439 \cdot X_1$$

(siehe Abschnitt 6.3) nicht als einfacher kausaler reglinearer Effekt von  $X_1$  auf  $Y_1$  interpretieren. Er kann lediglich als Durchschnitt der beiden einfachen kausalen reglinearen Effekte 0.4 und 0.0878 aus den Gleichungen 7.3.8 und 7.3.9 bezeichnet werden, denn bei Einschränkung des Geltungsbereichs unserer Aussage über eine einfache kausale reglineare Abhängigkeit der Variablen  $Y_1$  von  $X_1$  auf jeweils eine der beiden Bedingungen A bzw.  $\bar{A}$ , spräche wohl nichts dagegen, die durch die Gleichungen 7.3.8 und 7.3.9 beschriebenen Abhängigkeiten kausal zu interpretieren.

Formal würde man die Einschränkung des Geltungsbereichs wie folgt formulieren:  $Y_1$  ist von  $X_1$  auf dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P_{\bar{A}})$  einfach kausal reglinear abhängig mit dem einfachen kausalen reglinearen Effekt **0.4**, und auf dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P_A)$  mit dem Effekt **0.0878**. In der Sprache der Experimentellen Psychologie ausgedrückt, gilt also die durch Gleichung 7.3.8 beschriebene einfache kausale reglineare Abhängigkeit nur für die Population der Münze I, und die durch Gleichung 7.3.9 beschriebene nur für die Population der Münze II.

Offensichtlich ist auch bei der Frage nach der Populationsvalidität lediglich gefragt, auf welchen Wahrscheinlichkeitsraum sich die Aussage über eine kausale stochastische Abhängigkeit bezieht. Der *Wahrscheinlichkeitsraum repräsentiert demnach sowohl die Situation als auch die Population, auf die sich eine Aussage über eine stochastische Abhängigkeit bezieht*. Durch eine möglichst genaue Beschreibung der Untersuchungssituation und der beteiligten Versuchspersonen, wie dies in der Experimentellen Psychologie üblich ist, wird der zugrunde gelegte Wahrscheinlichkeitsraum gekennzeichnet.

## 7.4 Vergleiche der externen Validität

Wir gehen nun auf das im Abschnitt 6 behandelte Beispiel zurück. Im Abschnitt 6.4 haben wir gezeigt, daß der Koeffizient  $\gamma_{11}$  des einfachen logitlinearen Modells

$$(7.4.1) \quad E(Y_1 | X_1) \stackrel{fs}{=} P(B | X_1) \stackrel{fs}{=} \frac{\exp(\gamma_{10} + \gamma_{11} X_1)}{1 + \exp(\gamma_{10} + \gamma_{11} X_1)} = \\ = \frac{\exp(0.8473 + 1.9756 \cdot X_1)}{1 + \exp(0.8473 + 1.9756 \cdot X_1)}$$

nicht mit dem Koeffizienten  $\beta_{11} = 2.1972$  aus Gleichung 6.2.5 identisch ist, mit welcher die Daten der Tabelle 6.2.1 erzeugt wurden. Wir zeigen nun, daß wir den Koeffizienten  $\beta_{11} = 2.1972$  dann erhalten, wenn wir die logitlineare Abhängigkeit jeweils unter einer der beiden Bedingungen

$$(7.4.2) \quad \bar{A} := \text{der zweite Schalter ist aus,}$$

und

$$(7.4.3) \quad A := \text{der zweite Schalter ist an,}$$

untersuchen, so wie wir dies auch für die reglineare Abhängigkeit gemacht haben. Im Gegensatz zum reglinearen Modell wird sich aber hier zeigen, daß der Koeffizient  $\beta_{11} = 2.1972$  sowohl für die Bedingung A als auch  $\bar{A}$  gilt.

Bezeichnen wir mit

$$(7.4.4) \quad B := \{\omega \in \Omega: Y_1(\omega) = 1\}$$

weiterhin das Ereignis, daß die Münze auf die Metallseite fällt, und legen wir den Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P_{\bar{A}})$  (siehe Abschnitt 7.2) zugrunde, der die Situationsklasse repräsentiert, in welcher der zweite Schalter aus ist, so gilt

$$(7.4.5) \quad E_{\bar{A}}(Y_1 | X_1) \stackrel{P_{\bar{A}}-fs}{=} P(B | X_1) \stackrel{P_{\bar{A}}-fs}{=} \frac{\exp[\beta_{10}(\bar{A}) + \beta_{11} X_1]}{1 + \exp[\beta_{10}(\bar{A}) + \beta_{11} X_1]}.$$

Zur Berechnung der beiden unbekannten Parameter bilden wir nun die bedingten Wahrscheinlichkeiten<sup>22)</sup>  $P_{\bar{A}}(B | X_1=0)$  und  $P_{\bar{A}}(B | X_1=1)$ , aus denen wir nach einigen algebraischen Umformungen

$$(7.4.6) \quad \beta_{10}(\bar{A}) = \ln \left[ \frac{P_{\bar{A}}(B | X_1=0)}{1 - P_{\bar{A}}(B | X_1=0)} \right] = \ln [0.5 / (1 - 0.5)] = 0.0$$

---

<sup>22)</sup> Siehe zur Schreibweise die Bemerkung zur Definition A.4.1.

und

$$(7.4.7) \quad \beta_{10}(\bar{A}) + \beta_{11} = \ln \left[ \frac{P_{\bar{A}}(B | X_1=1)}{1 - P_{\bar{A}}(B | X_1=1)} \right] = \\ = \ln [0.9/(1 - 0.9)] = 2.1972$$

erhalten. Daraus ergibt sich der Koeffizient  $\beta_{11} = 2.1972$ , der mit  $\beta_{11} = 2.1972$  aus Gleichung 6.2.5 identisch ist.

Legen wir dagegen den Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P_A)$  zugrunde, so finden wir auf die analoge Weise für die Gleichung

$$(7.4.8) \quad E_A(Y_1 | X_1) \underset{P_{\bar{A}} \neq s}{=} P_A(B | X_1) \underset{P_A \neq s}{=} \frac{\exp[\beta_{10}(A) + \beta_{11} X_1]}{1 + \exp[\beta_{10}(A) + \beta_{11} X_1]},$$

die Parameter

$$(7.4.9) \quad \beta_{10}(A) = \ln \left[ \frac{P_A(B | X_1=0)}{1 - P_A(B | X_1=0)} \right] = \ln [0.9/(1 - 0.9)] = 2.1972$$

und

$$(7.4.10) \quad \beta_{10}(A) + \beta_{11} = \ln \left[ \frac{P_A(B | X_1=1)}{1 - P_A(B | X_1=1)} \right] = \\ = \ln [0.9878/(1 - 0.9878)] = 4.3940.$$

Daraus ergibt sich wieder der Koeffizient  $\beta_{11} = 2.1968$ , der bis auf Rundungsfehler mit  $\beta_{11} = 2.1972$  aus Gleichung 6.2.5 identisch ist. Dieser Koeffizient gilt also sowohl für den Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P_A)$  als auch für  $(\Omega, \mathcal{A}, P_{\bar{A}})$ , d.h. er gilt sowohl unter der Bedingung A, daß der zweite Schalter an ist, als auch unter der Bedingung  $\bar{A}$ , daß er aus ist. Hierin unterscheidet sich das logitlineare vom reglinearen Modell, denn die reglinearen Koeffizienten sind unter beiden Bedingungen verschieden, wie wir im Abschnitt 7.2 festgestellt haben. Zu dem entsprechenden Ergebnis würden wir auch gelangen, wenn wir anstelle des Beispiels mit zwei verschiedenen Situationen das im letzten Abschnitt behandelte Beispiel mit zwei verschiedenen Populationen betrachten würden.

Wie wir oben gezeigt haben, ist für das Beispiel des Abschnitt 6 die logitlineare Abhängigkeit der Münzwurfvariablen  $Y_1$  von der Magnetvariablen  $X_1$  mit dem Koeffizienten  $\beta_{11} = 2.1968$  sowohl in dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P_A)$  als auch  $(\Omega, \mathcal{A}, P_{\bar{A}})$  gültig, d.h. unter der Bedingung, daß der zweite Schalter an bzw. aus ist. Wir können also sagen, daß das logitlineare Modell in diesem speziellen Beispiel einen größeren Geltungsbereich, oder in der Sprache Campbells, eine größere externe Validität hat als das reglineare Modell.

## 7.5 Zusammenfassende Bemerkungen

Es wurde anhand des im letzten Kapitel dargestellten Beispiels erläutert, wie sich die Frage nach dem Geltungsbereich oder der externen Validität in der in diesem Beitrag vorgestellten Theorie einfacher kausaler reglinearer Abhängigkeit behandeln läßt. Es wurde gezeigt, daß sowohl die Frage nach der *Situationsvalidität*, als auch die nach der *Populationsvalidität*, die Frage nach dem zugrunde gelegten Wahrscheinlichkeitsraum ist.

Mit der in der Experimentellen Psychologie üblichen genauen Beschreibung der Untersuchungssituationen und der herangezogenen Versuchspersonen wird der zugrunde gelegte Wahrscheinlichkeitsraum charakterisiert. Verallgemeinerungen der in einem Wahrscheinlichkeitsraum gültigen Beziehungen auf einen anderen Wahrscheinlichkeitsraum können zwar hypothetisch vorgenommen werden, aber im allgemeinen gibt es für solche Verallgemeinerungen keine logischen oder mathematischen Begründungen. Daher muß jeweils eine empirische Überprüfung vorgenommen werden.

Es wurde gezeigt, daß in dem hier behandelten speziellen Beispiel das logitlineare Modell einen größeren Geltungsbereich oder eine größere ‚externe Validität‘ hat, als das reglineare Modell. Damit ist gemeint, daß in der Menge der Wahrscheinlichkeitsräume, in denen das logitlineare Modell gilt, die Menge der Wahrscheinlichkeitsräume, in denen das reglineare Modell gilt, enthalten ist.

## 8. Ausblick

### 8.1 Mehrvariablenmodelle

Die zentrale Fragestellung in diesem Beitrag war, unter welchen Voraussetzungen und in welchem Sinn sich die Koeffizienten einer einfachen Regressionsgleichung kausal interpretieren lassen. Obwohl diese Frage zunächst sehr speziell erscheinen mag, ist sie jedoch von recht weitreichender Bedeutung für die psychologische Forschung und ihr benachbarte Disziplinen. *Erstens* kann man diese Frage nämlich analog auch für andere Arten stochastischer Abhängigkeiten untersuchen. *Zweitens* läßt sich eine multiple Regressionsgleichung

$$(8.1.1) \quad E(Y | Z_m, m \in M) \stackrel{fs}{=} \alpha_{Y0} + \sum_{m \in M} \alpha_{Ym} Z_m$$

in mehrere einfache Regressionsgleichungen

$$(8.1.2) \quad E_B(Y | Z_m) \stackrel{p_B-fs}{=} \alpha_{Y0}(B) + \alpha_{Ym} Z_m, \quad m \in M,$$

auflösen, wenn B das Ereignis bezeichnet, daß die anderen Variablen aus der

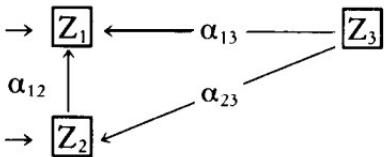


Abb. 8.1.1: Pfeildiagramm eines einseitig kausalen reglinearen Modells mit drei Variablen.

Familie ( $Z_m$ ,  $m \in M$ ) konstant sind. Auf diesem Sachverhalt baut die Theorie der direkten und auch der totalen kausalen reglinearen Abhangigkeit auf, die ausfuhrlich von Steyer (1982) behandelt wird.

*Drittens* enthalten eine ganze Reihe von Modellklassen als Kern solche Regressionsgleichungen. Neben der eigentlichen Regressionsanalyse ist da an erster Stelle die *Varianzanalyse* zu nennen, bei der die Variablen  $Z_m$ ,  $m \in M$ , die Zugehorigkeit zu bestimmten (z.B. experimentellen) Gruppen anzeigen. Aber auch die *Faktorenanalyse* basiert, auer auf der Annahme der bedingten stochastischen Unabhangigkeit, auf Regressionsgleichungen fur jede der beobachtbaren oder manifesten Variablen  $Y_r$ ,  $r \in R$ , unter den Faktoren oder latenten Variablen  $Z_m$ ,  $m \in M$ . Schlielich kann auch ein rekursives *Strukturgleichungsmodell* als ein System von solchen Regressionsgleichungen aufgefat werden, wobei ein Teil der Gleichungen die Beziehungen zwischen den manifesten und den latenten Variablen spezifiziert und ein anderer Teil der Gleichungen die Beziehungen der latenten Variablen untereinander.

Diese Verfahren, und damit die darin enthaltenen Regressionsgleichungen, konnen zum einen zur *Beschreibung* statistischer Zusammenhange verwendet werden, sie konnen zum anderen aber auch Bestandteile von Theorien zur kausalen **Erklarung** statistischer Zusammenhange sein. So besteht beispiels-

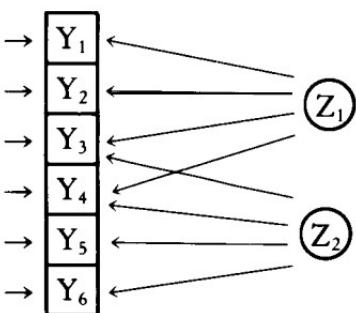


Abb. 8.1.2: Pfeilschema eines kausalen Modells mit zwei latenten Variablen (Faktoren)  $Z_1$  und  $Z_2$ , welche die stochastischen Abhangigkeiten zwischen den manifesten Variablen  $Y_1$  bis  $Y_6$  erklaren.

weise das pfadanalytische Modell der Abbildung 8.1.1 aus den beiden Regressionsgleichungen

$$(8.1.3) \quad E(Z_1 | Z_2, Z_3) \underset{\text{fs}}{=} \alpha_{10} + \alpha_{12} Z_2 + \alpha_{13} Z_3$$

und

$$(8.1.4) \quad E(Z_2 | Z_3) \underset{\text{fs}}{=} \alpha_{20} + \alpha_{23} Z_3,$$

das faktorenanalytische Modell der Abbildung 8.1.2 aus den sechs Gleichungen

$$(8.1.5) \quad E(Y_r | Z_1, Z_2) \underset{\text{fs}}{=} \beta_{r0} + \beta_{r1} Z_1 + \beta_{r2} Z_2, \quad r \in \{1, \dots, 6\},$$

und das Strukturgleichungsmodell der Abbildung 8.1.3 außer aus den Gleichungen 8.1.3 bis 8.1.4 auch aus den Gleichungen

$$(8.1.6) \quad E(X_s | Z_1, Z_2, Z_3) \underset{\text{fs}}{=} \gamma_{s0} + \gamma_{s3} Z_3, \quad s \in \{1, 2, 3\},$$

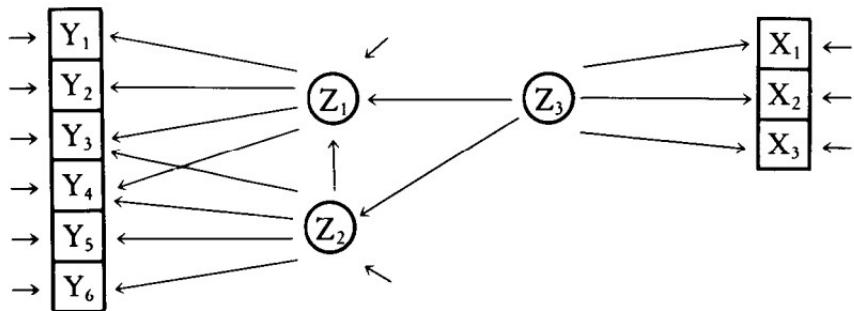


Abb. 8.1.3: Pfeilschema eines kausalen Modells mit drei latenten Variablen  $Z_1$ ,  $Z_2$  und  $Z_3$ , die zum einen die manifesten Variablen  $Y_1$  bis  $Y_6$  bzw.  $X_1$  bis  $X_3$  beeinflussen, zum anderen aber auch untereinander kausal abhängig sind.

und

$$(8.1.7) \quad E(Y_r | Z_1, Z_2, Z_3) \underset{\text{fs}}{=} \beta_{r0} + \beta_{r1} Z_1 + \beta_{r2} Z_2, \quad r \in \{1, \dots, 6\}.$$

Liegen, wie bei den oben genannten Modellen, mehrere unabhängige Variablen vor, so kann man die Frage nach einer *direkten* kausalen reglinearen Abhängigkeit stellen, eine Frage, die für die psychologische Theorienbildung genauso bedeutsam ist, wie die nach einfacher kausaler reglinearer Abhängigkeit. Wird beispielsweise behauptet, daß

$$(8.1.8) \quad Y := \text{absolute IQ-Differenz von Zwillingen}$$

einfach kausal reglinear abhängig ist von

$$(8.1.9) \quad x := \begin{cases} 1, & \text{wenn die Zwillinge eineig,} \\ -1, & \text{wenn sie zweieiig sind,} \end{cases}$$

so ist es nicht nur wichtig zu untersuchen, ob diese Aussage an sich richtig ist, ob es also potentielle Störvariablen W gibt, die den Regressionskoeffizienten  $\alpha_{YX}$  der Gleichung

$$(8.1.10) \quad E(Y | X) \underset{fs}{=} \alpha_{Y0} + \alpha_{YX} X$$

modifizieren, wenn eine solche Variable W konstant gehalten wird; vielmehr ist es auch von Bedeutung zu untersuchen, durch welche Variablen eine solche einfache kausale reglineare Abhängigkeit vermittelt wird. So wäre es prinzipiell durchaus denkbar, daß diese Abhängigkeit nur über die zwischen X und Y anzuordnende Variable

$$(8.1.11) \quad Z := \text{Ähnlichkeit des elterlichen Erziehungsverhaltens}$$

zustande kommt, d.h. daß die bezüglich {X,Y,Z} **direkte** kausale reglineare Abhängigkeit der Variablen Y von X Null wäre. Dies wäre durchaus kein Widerspruch zur Aussage, daß Y von X *einfach* kausal reglinear abhängig ist (vgl. zur Diskussion dieses Beispiels Eysenck, 1979, S. 105).

Die Grundidee direkter kausaler reglinearer Abhängigkeit einer stochastischen Variablen Y von einer stochastischen Variablen X - direkt bezüglich einer gegebenen Familie von Variablen - ist, daß eine einfache kausale reglineare Abhängigkeit Y von X besteht, und zwar *bei Konstanz aller anderen Variablen in dieser Familie, die Y vorgeordnet sind*, und daß der Steigungskoeffizient  $\alpha_{YX}$  über alle Ausprägungen der konstantgehaltenen Variablen hinweg der gleiche bleibt.

Diese Definition ermöglicht, daß die Koeffizienten einer multiplen Regressionsgleichung als direkte kausale reglineare Effekte interpretiert werden können, wenn die in der Definition genannten Bedingungen erfüllt sind. Damit wird es nicht nur möglich, die Direktheit einer Abhängigkeit zu überprüfen, sondern auch potentielle Störvariablen statistisch zu kontrollieren oder „statistisch konstant zu halten“.

Falsifiziert wäre eine Aussage, daß Y von jeder Variablen  $Z_m$ ,  $m \in M$ , mit dem Effekt  $\alpha_{Ym}$  direkt kausal reglinear abhängig ist (siehe Gleichung 8.1.1), wenn für eine potentielle Störvariable W, gezeigt werden kann, daß die Gleichung

$$(8.1.12) \quad E(Y | W, Z_m, m \in M) \underset{fs}{=} \sum_{m \in M} \alpha_{Ym} Z_m + f(W)$$

nicht gilt, wobei f(W) eine (meßbare) Funktion von W ist. Beispieldweise wäre  $\alpha_{Y0} + \alpha_{YW} \cdot W$  eine solche Funktion von W,  $\alpha_{Y0} + \alpha \cdot Z_m \cdot W$  hingegen nicht, wenn  $\alpha \neq 0$ . Dabei ist zu beachten, daß die Koeffizienten  $\alpha_{Ym}$  in den Glei-

chungen 8.1.1 und 8.1.12 identisch sind. Andernfalls wäre die oben genannte Aussage falsifiziert. Für eine ausführlichere und vollständigere Behandlung der Theorie direkter kausaler reglinearer Abhängigkeit sei auf Steyer (1982) verwiesen.

## 8.2 Beschreibende und erklärende reglineare Modelle

Bei der Beantwortung der Frage nach dem formalen Unterschied zwischen einem beschreibenden und einem erklärenden reglinearen Modell, und damit nach der Bedeutung einer kausalen Interpretation von Regressionsparametern sind wir von einem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  ausgegangen, der in Anwendungsfällen eine bestimmte Untersuchungssituationsklasse oder ein bestimmtes Experiment beschreibt, und zwar dadurch, daß jedem einfachen und verbundenen Ereignis  $A \in \mathcal{A}$  eine bestimmte feste Wahrscheinlichkeit  $P(A)$  zugeordnet ist. Auf einem solchen Wahrscheinlichkeitsraum sind die betrachteten stochastischen Variablen definiert, deren Abhängigkeiten voneinander möglicherweise durch Regressionsgleichungen beschreibbar sind. Bis zu diesem Punkt besteht zwischen einem beschreibenden und einem erklärenden Modell kein Unterschied.

Damit die durch eine einfache Regressionsgleichung

$$(8.2.1) \quad E(Y | X) = \alpha_{Y0} + \alpha_{YX} X$$

beschriebene stochastische Abhängigkeit zwischen der Variablen  $X$  und einer Variablen  $Y$  unkonfundiert ist und daher bei entsprechender zeitlicher Geordnetheit der Variablen kausal interpretiert werden kann, müssen zwei Bedingungen erfüllt sein, die rein formal definierbar sind. Verbal können diese Bedingungen folgendermaßen beschrieben werden: *Vorgeordnetheit*: Die Variable  $X$  muß der Variablen  $Y$  vorgeordnet sein. *Invarianz*: Der Regressionskoeffizient  $\alpha_{YX}$  wird nicht durch potentielle Störvariablen modifiziert. Diese Vorstellung wurde als eine bestimmte Eigenschaft der gemeinsamen Verteilung der stochastischen Variablen auf dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  präzisiert. Wie die genannten zwei Bedingungen formaler und damit genauer definiert sind, wie die zweite in Anwendungsfällen empirisch überprüft werden kann und wie empirische Untersuchungen so anzulegen sind, daß diese Bedingungen nach Möglichkeit erfüllt ist, wurde in diesem Beitrag für den Fall einer einzigen unabhängigen Variablen ausführlich beschrieben und am Ende der einzelnen Abschnitte jeweils zusammengefaßt. Eine explizite Darstellung für mehrere unabhängige Variablen findet man in Steyer (1982). Als wichtigste Ergebnisse können die folgenden Punkte hervorgehoben werden:

1. Es wurde gezeigt, daß sich die *experimentellen Kontrolltechniken der Randomisierung, Parallelisierung und Konstanthaltung* auf eine rein formale Theo-

rie basieren lassen. Bisher beruhten die Kontrolltechniken auf Plausibilitätsüberlegungen wie etwa der folgenden, die auf Mills Unterschiedsmethode zurückgeht (Mill, 1885, S. 91 ff.): Wenn man davon ausgeht, daß ein Sachverhalt W von vielen anderen möglicherweise unbekannten Sachverhalten A,B,C,D,. . . verursacht wird, und man untersuchen will, ob ein bestimmter Sachverhalt C zu diesen Ursachen zählt, so muß man Beobachtungen unter zwei Bedingungskombinationen anstellen, die sich lediglich durch C unterscheiden, also A,B,C,D,. . . und A,B,C,D,. . . . Tritt dann nur in der ersten Bedingungskombination die Wirkung W auf, dann kann man davon ausgehen, daß C Ursache für W ist. Der Herstellung der sonst gleichen Bedingungen A,B,. . . dienen die oben genannten Kontrolltechniken.

In gewisser Hinsicht kann man die hier vorgestellte Theorie als eine Verfeinerung der Millschen Theorie<sup>23)</sup> ansehen, in dem Sinne, daß auch nichtdeterministische kausale Abhängigkeiten zugelassen werden. Mit dem hier behandelten stochastischen Begriff kausaler Abhängigkeit kann C auch dann als Ursache von W angesehen werden, wenn C nicht immer zusammen mit W auftritt, auch nicht in einer streng kontrollierten Situation, sondern nur mit erhöhter Wahrscheinlichkeit.

2. Es wurde gezeigt, daß *Kausalität nicht zwangsläufig Determinismus* bedeutet. Beim Beispiel ‚Münzen und Elektromagnet‘ beeinflußt der Zustand des Elektromagneten die Ergebnisse des Münzwerfens nur stochastisch in dem Sinn, daß bei angeschaltetem Elektromagneten die Wahrscheinlichkeit, daß die Münzen auf die Metallseite fallen, erhöht ist. Um die Abhängigkeit der Münzwurfergebnisse vom Zustand des Elektromagneten als kausal zu bezeichnen, muß diese also keineswegs deterministisch sein. Selbst wenn die Erhöhung der Wahrscheinlichkeit nur minimal ist, z.B. von 0.5 auf 0.51, wäre dies ein hinreichender Grund, von einer kausalen Abhängigkeit zu sprechen, deren praktische Bedeutsamkeit allerdings geringer ist, als wenn die Wahrscheinlichkeit von 0.5 auf 0.9 erhöht wird, wie in dem in Abschnitt 2 dargestellten Beispiel. Die Stärke der Abhängigkeit und die Frage, ob die Abhängigkeit kausal ist, sind also zwei völlig verschiedene Fragen.

3. Die in der Experimentellen Psychologie und Pädagogik entwickelte Vorstellung ‚interner Validität‘ wurde zum rein formalen Begriff einfacher kausaler reglinearer Abhängigkeit präzisiert. Die formale Definition ermöglicht die Ableitung von Konsequenzen, die in Anwendungsfällen *empirisch überprüfbar und gegebenenfalls falsifizierbar* sind. Diese grundlegende Forderung an eine empirische wissenschaftliche Theorie ist wohl bei der bisherigen pfadana-

---

<sup>23)</sup> Eine zusammenfassende Darstellung der Millschen Theorie findet man bei Diekoff (1981, S. 24ff.), der auch über die Beiträge von Hume, Kant, Popper u.a. zum Thema kausale Abhängigkeit referiert.

lytischen Vorgehensweise nicht hinreichend erfüllt. Zwar werden Modelltests zur Überprüfung der Frage durchgeführt, ob das statistische Modell mit der jeweils vorliegenden Kovarianzmatrix verträglich ist. Die wesentliche Frage nach der Kausalität der mit dem Modell beschriebenen Beziehungen wird aber nicht überprüft. Die Verträglichkeit eines Strukturgleichungsmodells mit der Kovarianzmatrix ist jedoch kein Hinweis dafür, daß die Strukturparameter kausal interpretiert werden können, sondern nur dafür, daß sich die betreffenden Daten mit diesem Modell *beschreiben* lassen. Kausalität wird bei einem Strukturgleichungsmodell bisher vorausgesetzt, etwa durch die Annahme der „Geschlossenheit“, aber nicht überprüft. Auch wurde in der bisherigen Literatur nie angegeben, wie die „Geschlossenheit“ prinzipiell falsifiziert werden kann.

4. Aus der hier vorgestellten Theorie lassen sich einige *Möglichkeiten zur Falsifizierung* einer Aussage ableiten, daß der Koeffizient aus der Gleichung 8.2.1 kausal interpretiert werden kann. So ist z.B. jede der beiden folgenden Aussagen in einem Anwendungsfall hinreichend für eine solche Falsifizierung:

a) Es besteht eine Interaktion im varianzanalytischen Sinn zwischen X und einer potentiellen Störvariablen W. Dabei ist es gleichgültig, ob diese erhoben wurde oder nicht. Wesentlich ist, daß sie in der betreffenden Situationsklasse prinzipiell erhoben werden könnte.

b) Der Koeffizient  $\beta_{YX}$  von X in der Gleichung

$$(8.2.2) \quad E(Y | X, W) \stackrel{f_s}{=} \beta_{Y0} + \beta_{YX} X + \beta_{YW} W$$

ist nicht identisch mit aus Gleichung 8.2.1, wobei W eine beliebige potentielle Störvariable ist, die in der gleichen Situationsklasse wie die anderen Variablen erhoben wird oder werden könnte. Bei der Erweiterung von  $E(Y | X)$  auf  $E(Y | X, W)$  muß also der Koeffizient unverändert bleiben, wenn er kausal interpretierbar sein soll.

Die Konsequenz einer Falsifikation aufgrund von Punkt a ist eine Einschränkung des Geltungsbereichs. Für jede Ausprägung der mit X interagierenden Variablen W gilt u.U. eine andere einfache kausale reglineare Abhängigkeit. Bei einer Falsifikation aufgrund von Punkt b gilt u.U. eine einfache kausale reglineare Abhängigkeit, wenn  $W = w$  konstant ist.

5. Prinzipiell ist es sowohl in experimentellen als auch in nichtexperimentellen Untersuchungen möglich, kausale Abhängigkeiten festzustellen, aber in keiner Art von Untersuchung kann man *beweisen*, daß eine kausale Abhängigkeit vorliegt. Die Möglichkeit zur *Falsifizierung* einer Aussage über eine kausale Abhängigkeit besteht jedoch in beiden Forschungssituationen. Diese unterscheiden sich nur darin, in welchem Ausmaß es möglich ist, durch versuchsplanerische Maßnahmen potentielle Störvariablen zu kontrollieren. Je geringer

die Möglichkeiten der aktiven Kontrolle potentieller Störvariablen sind, desto wichtiger wird es, potentielle Störvariablen mitzuerheben und diese entweder mit in das Modell einzubeziehen, oder die unter Punkt 4 genannten ‚Kausalitätstests‘ durchzuführen.

## Weiterführende Literatur

Die Verallgemeinerung der hier dargestellten Theorie einfacher kausaler reglinearer Abhängigkeit auf Mehrvariablenmodelle findet man bei Steyer (1982). Eine weitere Verallgemeinerung auf nichtrekursive Modelle, in denen auch wechselseitige Abhängigkeiten vorkommen können, ist in Vorbereitung. Eine Übersicht über Stichprobenprobleme wie Parameterschätzung und Hypothesen- bzw. Modellbewertungsverfahren insbesondere für komplexe Modelle gibt Bentler (1980), der auch verschiedene Rechenprogramme für diese Zwecke angibt. Für die einfacheren Varianz- und regressionsanalytischen Modelle siehe auch die entsprechenden Beiträge in diesem Band. Ein Überblick über verschiedene Schätzverfahren gibt Klein (1978). Zur Einführung in pfadanalytische Modelle sei auch auf Asher (1976) und Kenny (1979) hingewiesen. Als weiterführende Arbeiten und Anwendungen seien Aigner und Goldberger (1977), Anderson und Evans (1974), Blalock (1962, 1964, 1969, 1971a,b), Blalock und Blalock (1968), Boudon (1963, 1965a,b, 1968, 1976a,b), Hummell und Ziegler (1976a) sowie Marsden (1982) genannt.

## Anhang

### A.1 Einleitende Bemerkungen

In diesem Anhang sind einige im Text häufig verwendete Definitionen und Theoreme der Wahrscheinlichkeitstheorie zusammengestellt. Damit soll lediglich die Angabe der mathematischen Grundlagen bei Ableitungen und Beweisen vereinfacht werden. Mit diesem Anhang ist weder ein Anspruch auf Vollständigkeit verbunden, noch ist intendiert, den Leser in die Wahrscheinlichkeitstheorie einzuführen. Dazu sei auf Bauer (1974), Breiman (1968), Gänssler und Stute (1977), Hinderer (1980) sowie Loève (1977, 1978) verwiesen. Zum Nachschlagen eignet sich Müller (1975).

### A.2 Erwartungswert

*Definition A.2.1.* Es sei  $X: \Omega \rightarrow \mathbb{R}$  eine stochastische Variable<sup>24)</sup> auf dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$ . Außerdem sei  $X$  bezüglich  $P$  integrierbar oder nichtnegativ. Genau dann, wenn

<sup>24)</sup>  $\mathbb{R}$  ist dabei die Menge der reellen Zahlen. Wir betrachten in diesem Beitrag nur reellwertige stochastische Variablen. Viele Autoren gebrauchen anstelle der Bezeichnung ‚stochastische Variable‘ die Bezeichnung ‚Zufallsvariable‘.

$$(A.2.1) \quad E(X) := \int X \, dP,$$

heißt  $E(X)$  der Erwartungswert von  $X$  (bezüglich des Wahrscheinlichkeitsmaßes  $P$ ).

*Theorem A.2.1.* Ist  $X$  eine diskrete stochastische Variable auf dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  mit dem Erwartungswert  $E(X)$  und den Werten  $x_i, i \in I$ , wobei  $I$  eine abzählbare Indexmenge ist, und bezeichnet

$$(A.2.2) \quad A_i := \{\omega \in \Omega : X(\omega) = x_i\}$$

das Ereignis, daß  $X$  den Wert  $x_i$  angenommen hat, dann gilt

$$(A.2.3) \quad E(X) = \sum_{i \in I} x_i P(A_i)$$

(vgl. z.B. Bauer, 1974, S. 138f.).

*Theorem A.2.2.* Sind  $X$  und  $Y$  stochastische Variablen auf dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  mit den Erwartungswerten  $E(X)$  bzw.  $E(Y)$  und sind  $\alpha$  und  $\beta$  reelle Zahlen, so gelten

$$(A.2.4) \quad E(\alpha) = \alpha \quad \text{und} \quad E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y)$$

(vgl. z.B. Bauer, 1974, S. 65).

### A.3 Varianz und Kovarianz

*Definition A.3.1.* Es seien  $X$  und  $Y$  stochastische Variablen auf dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  mit den Erwartungswerten  $E(X)$  bzw.  $E(Y)$ . Genau dann, wenn

$$(A.3.1) \quad C(X, Y) := E([X - E(X)] \cdot [Y - E(Y)]),$$

heißt  $C(X, Y)$  *Kovarianz* von  $X$  und  $Y$ . Genau dann, wenn

$$(A.3.2) \quad V(X) := C(X, X) = E([X - E(X)]^2),$$

heißt  $V(X)$  *Varianz* von  $X$  (vgl. z.B. Gänssler & Stute, 1977, S. 8lf.).

*Theorem A.3.1.* Sind  $X$  und  $Y$  stochastische Variablen auf dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  mit den Erwartungswerten  $E(X)$  bzw.  $E(Y)$ , so gilt<sup>25)</sup>

$$(A.3.3) \quad C(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y)$$

und, falls  $\alpha$  und  $\beta$  reelle Zahlen sind, auch

$$(A.3.4) \quad C(\alpha X, \beta Y) = \alpha C(X, Y) \beta.$$

*Theorem A.3.2.* Sind  $W, X, Y$  und  $Z$  stochastische Variablen auf dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$ , mit den Erwartungswerten  $E(W), E(X), E(Y)$  bzw.  $E(Z)$  und sind  $\alpha, \beta, \gamma$  und  $\delta$  reelle Zahlen, so gilt<sup>25)</sup>

$$(A.3.5) \quad C[(\alpha W + \beta X), (\gamma Y + \delta Z)] = \\ = \alpha C(W, Y) \gamma + \alpha C(W, Z) \delta + \beta C(X, Y) \gamma + \beta C(X, Z) \delta.$$

## A.4 Bedingter Erwartungswert

*Definition A.4.1.* Es sei  $Y$  eine stochastische Variable auf dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$ ,  $B \in \mathcal{A}$  ein Ereignis mit  $P(B) > 0$ , und  $1_B$  sei die Indikatorvariable<sup>26)</sup> von  $B$ . Genau dann, wenn

$$(A.4.1) \quad E(Y | B) := \frac{1}{P(B)} E(1_B \cdot Y),$$

heißt  $E(Y | B)$  *bedingter Erwartungswert von  $Y$  gegeben das Ereignis  $B$* . Genau dann, wenn  $Y$  dabei nur die Werte 0 und 1 annehmen kann, und

$$(A.4.2) \quad A := \{\omega \in \Omega : Y(\omega) = 1\}$$

das Ereignis ist, daß  $Y$  den Wert 1 angenommen hat, heißt

$$(A.4.3) \quad P(A | B) := E(Y | B) = P(A \cap B) / P(B)$$

die *bedingte Wahrscheinlichkeit von  $A$  gegeben das Ereignis  $B$*  (vgl. z.B. Bauer, 1974, s. 290ff.).

*Bemerkung.* Mit  $E(Y | B)$  und  $P(A | B)$  äquivalente Schreibweisen sind  $E_Y$  bzw.  $P_A$ . Mit dieser Notation wird hervorgehoben, daß  $E_B(Y)$  der Erwartungswert von  $Y$  bezüglich des bedingten Wahrscheinlichkeitsmaßes  $P_B$  bzw. daß  $P_S(A)$  der Wert des bedingten Wahrscheinlichkeitsmaßes  $P_B$  für das Ereignis  $A$  ist. Das Maß  $P_B$  ist durch Gleichung A.4.3 definiert, die dabei auf jedes  $A \in \mathcal{A}$  anzuwenden ist.

*Theorem A.4.1.* Es seien  $X$  und  $Y$  stochastische Variablen auf dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  mit den Erwartungswerten  $E(X)$  bzw.  $E(Y)$  sowie  $\alpha$  und  $\beta$  reelle Zahlen. Wenn  $E(Y | B)$  der bedingte Erwartungswert von  $Y$  gegeben das Ereignis  $B \in \mathcal{A}$  ist, dann gelten die folgenden Gleichungen<sup>27)</sup>:

<sup>25)</sup> Die in Theorem A.3.1 und A.3.2 behaupteten Eigenschaften folgen aus der Definition A.3.1 und Theorem A.2.2.

<sup>26)</sup> Die stochastische Variable  $1_B$  auf dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  heißt Indikatorvariable für das Ereignis  $B \in \mathcal{A}$  und genau dann, wenn  $1_B(\omega) = 1$ , falls  $\omega \in B$  und  $1_B(\omega) = 0$  andernfalls.

<sup>27)</sup> Die in Theorem A.4.1 behaupteten Eigenschaften folgen direkt aus der Definition A.4.1 und Theorem A.3.1.

$$(A.4.4) \quad E(Y | \Omega) = E(Y),$$

$$(A.4.5) \quad E(\alpha | B) = \alpha,$$

$$(A.4.6) \quad E(\alpha X + \beta Y | B) = \alpha E(X | B) + \beta E(Y | B).$$

## A.5 Bedingte Erwartung

*Definition A.5.2.* Es sei  $Y$  eine stochastische Variable auf dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  mit Erwartungswert  $E(Y)$ ,  $\mathcal{B}$  sei eine Teilsigmaalgebra von  $\mathcal{A}$ , d.h.  $\mathcal{B} \subset \mathcal{A}$ , und  $l_B$  sei die Indikatorvariable für  $B \in \mathcal{B}$ . Genau dann, wenn

$$(A.5.1) \quad E[l_B \cdot E(Y | \mathcal{B})] = E(l_B \cdot Y), \quad \text{für alle } B \in \mathcal{B},$$

und  $E(Y | \mathcal{B})$  messbar bezüglich  $\mathcal{B}$  ist, heißt  $E(Y | \mathcal{B})$  bedingte Erwartung von  $Y$  unter **der Sigmaalgebra  $\mathcal{B}$**  (vgl. z.B. Gänssler & Stute, 1977, S. 187).

Genau dann, wenn  $Y = l_A$ , heißt  $E(Y | \mathcal{B})$  bedingte Wahrscheinlichkeit von  $A$  unter der Sigmaalgebra  $\mathcal{B}$  und wird dann auch mit  $P(A | \mathcal{B})$  notiert (vgl. z.B. Bauer, 1974, S. 296)<sup>28)</sup>.

*Bemerkungen.* Durch diese Definition ist die bedingte Erwartung  $E(Y | \mathcal{B})$  nur fast sicher eindeutig bestimmt. Man spricht daher auch von verschiedenen Versionen der bedingten Erwartung, die jedoch fast sicher gleich sind, was wir mit dem Zeichen „ $\stackrel{P_A}{=}$ “ abkürzen. Aussagen über bedingte Erwartungen gelten daher im allgemeinen nur fast sicher. Ist die bedingte Erwartung bezüglich eines anderen Wahrscheinlichkeitsmaßes als  $P$  gemeint, z.B. bezüglich  $P_A$ , dann schreiben wir entsprechend  $E_A(Y | \mathcal{B})$ . Aussagen über eine solche bedingte Erwartung gelten entsprechend nur  $P_A$ -fast sicher, und verschiedene Versionen sind  $P_A$ -fast sicher gleich, was wir durch „ $\stackrel{P_A}{=}$ “ abkürzen. Ist die Sigmaalgebra  $\mathcal{B}$  von einer stochastischen Variablen  $X$  auf  $(\Omega, \mathcal{A}, P)$  erzeugt, so schreiben wir anstatt  $E(Y | \mathcal{B})$  auch  $E(Y | X)$ . Ist  $\mathcal{B}$  von einer Familie  $(X_i, i \in I)$  von stochastischen Variablen erzeugt, so notieren wir die bedingte Erwartung  $E(Y | \mathcal{B})$  auch mit  $E(Y | X_i, i \in I)$ . Ist die Indexmenge  $I$  abzählbar, d.h. ist  $I = \{1, \dots, Q\}$ ,  $Q \in \mathbb{N}$ , so schreiben wir auch  $E(Y | X_1, \dots, X_Q)$  (vgl. auch Bauer, 1974, S. 292).

*Theorem A.5.1.* Wenn  $E(Y | \mathcal{B})$  die bedingte Erwartung der stochastischen Variablen  $Y$  auf dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  unter der Sigmaalgebra  $\mathcal{B}$  ist, dann gelten die folgenden Aussagen:

$$(A.5.2) \quad E[E(Y | \mathcal{B}) | B] = E(Y | B), \quad \text{für alle } B \in \mathcal{B}, \text{ mit } P(B) > 0,$$

$$(A.5.3) \quad E[E(Y | \mathcal{B})] = E(Y),$$

---

<sup>28)</sup> Man beachte den Unterschied zwischen der bedingten Wahrscheinlichkeit  $P(A | B)$  von  $A$  gegeben das Ereignis  $B$  und der bedingten Wahrscheinlichkeit  $P(A | \mathcal{B})$  von  $A$  unter der Sigmaalgebra  $\mathcal{B}$ .

$$(A.5.4) \quad E[Y - E(Y | \mathcal{B})] = 0,$$

$$(A.5.5) \quad E(Y | \mathcal{B}) \stackrel{\text{fs}}{=} Y, \quad \text{falls } Y \text{ messbar bez\"uglich } \mathcal{B} \text{ ist,}$$

$$(A.5.6) \quad E(Y | \mathcal{B}) \stackrel{\text{fs}}{=} E(X | \mathcal{B}), \quad \text{falls } X \stackrel{\text{fs}}{=} Y$$

(vgl. z.B. Bauer, 1974, S. 293)

*Theorem A.5.2.* Wenn  $E(Y | \mathcal{B})$  die bedingte Erwartung der stochastischen Variablen  $Y$  auf dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  unter der Sigmaalgebra  $\mathcal{B}$  ist, und  $\mathcal{C}$  ist ein Teilsigmaalgebra von  $\mathcal{B}$ , d.h.  $\mathcal{C} \subset \mathcal{B}$ , dann gelten folgende Aussagen:

$$(A.5.7) \quad E[E(Y | \mathcal{B}) | \mathcal{C}] \stackrel{\text{fs}}{=} E[E(Y | \mathcal{C}) | \mathcal{B}] \stackrel{\text{fs}}{=} E(Y | \mathcal{C}),$$

und

$$(A.5.8) \quad E[Y - E(Y | \mathcal{B}) | \mathcal{C}] \stackrel{\text{fs}}{=} 0$$

(vgl. z.B. Bauer, 1974, S. 293).

*Theorem A.5.3.* Wenn  $X$  und  $Y$  stochastische Variablen auf dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  sind mit den Erwartungswerten  $E(X)$  bzw.  $E(Y)$  sowie  $\alpha$  und  $\beta$  reelle Zahlen, dann gelten die folgenden Aussagen:

$$(A.5.9) \quad E(\alpha | \mathcal{B}) \stackrel{\text{fs}}{=} \alpha,$$

$$(A.5.10) \quad E(\alpha X + \beta Y | \mathcal{B}) \stackrel{\text{fs}}{=} \alpha E(X | \mathcal{B}) + \beta E(Y | \mathcal{B}),$$

$$(A.5.11) \quad E(X \cdot Y | \mathcal{B}) \stackrel{\text{fs}}{=} X \cdot E(Y | \mathcal{B}), \quad \text{falls } X \text{ bez\"uglich } \mathcal{B} \text{ messbar ist}$$

(vgl. z.B. Bauer, 1974, S. 293).

*Bemerkung.*  $X$  ist z.B. dann bez\"uglich  $\mathcal{B}$  messbar, wenn  $\mathcal{B}$  die von  $(X_q, q \in Q)$  erzeugte Sigmaalgebra ist, und  $X = X_q$ , f\"ur ein  $q \in Q$ .

*Theorem A.5.4.* Wenn  $X$  eine bez\"uglich  $\mathcal{B}$  messbare stochastische Variable auf dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  ist, und es gilt f\"ur die bedingte Erwartung  $E(Y | \mathcal{B})$  von  $Y$  unter  $\mathcal{B}$

$$(A.5.12) \quad E(Y | \mathcal{B}) \stackrel{\text{fs}}{=} E(Y),$$

dann folgt

$$(A.5.13) \quad C(X, Y) = 0.$$

*Beweis.* Nach Theorem A.3.1 gilt

$$(A.5.14) \quad C(X, Y) = E(X \cdot Y) - E(X) E(Y).$$

Wenn wir Gleichung A.5.3 auf  $X \cdot Y$  anwenden, erhalten wir daraus

$$(A.5.15) \quad C(X,Y) = E[E(X \cdot Y | \mathcal{B})] - E(X) \cdot E(Y).$$

Nach Gleichung A.5.11 folgt dann

$$(A.5.16) \quad C(X,Y) = E[X \cdot E(Y | \mathcal{B})] - E(X) \cdot E(Y),$$

und nach Gleichung A.5.12 und A.2.4

$$(A.5.17) \quad C(X,Y) = E(X) \cdot E(Y) - E(X) \cdot E(Y).$$

*Theorem A.5.5.* Es seien X und Y stochastische Variablen auf dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  und  $E(Y)$  sei der Erwartungswert von Y. Wenn X und Y stochastisch unabhängig sind, dann gilt

$$(A.5.18) \quad E(Y | X) \stackrel{\text{fs}}{=} E(Y)$$

(vgl. Breiman, 1968, S. 74).

## Literatur

- Ahrens, H. 1968. Varianzanalyse. Berlin: Akademie-Verlag.
- Aigner, D. J. & Goldberger, A. S. (Hrsg.). 1977. Latent variables in socio-economic models. Amsterdam: North-Holland.
- Anderson, J. G. & Evans, F. B. 1974. Causal models in educational research: Recursive models. American Educational Research Journal, 11, 29-39.
- Anderson, T. W. 1954. On estimation of parameters in latent structure analysis. Psychometrika, 19, 1-10.
- Anderson, T. W. 1958. An introduction to multivariate statistical analysis. New York: Wiley.
- Anderson, T. W. 1959. Some scaling models and estimation procedures in the latent class model. In: Grenander, O. Probability and statistics. The Harold Cramer Volume. New York: Wiley, 9-38.
- Anderson, T. W. & Rubin, H. 1956. Statistical inference in factor analysis. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 5, 111-150.
- Asher, H. B. 1976. Causal modeling. Beverly Hills: Sage.
- Bagozzi, R. P. 1980. Causal models in marketing. New York: Wiley.
- Bartenwerfer, H. & Raatz, U. 1979. Methoden der Psychologie. Wiesbaden: Akademische Verlagsgesellschaft.
- Bauer, H. 1974. Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie. Berlin: de Gruyter.

- Bentler, P. M. 1980. Multivariate analysis with latent variables: Causal modeling. *Annual Review of Psychology*, 31, 419-456.
- Bishop, Y. Y. M., Fienberg, S. E. & Holland, P. W. 1975. Discrete multivariate analysis: Theory and Practice. Cambridge: MIT Press.
- Blalock, H. M. Jr. 1962. Further observations on asymmetric causal models. *American Sociological Review*, 27, 542-545.
- Blalock, H. M. Jr. 1964. Causal inferences in nonexperimental research. Chapel Hill: University of North Carolina Press.
- Blalock, H. M. Jr. 1969. Theory construction. From verbal to mathematical formulations. Englewood-Cliffs, N. J.: Prentice-Hall.
- Blalock, H. M. Jr. 1971a. (Hrsg.). Causal models in the social sciences. Chicago: Aldine-Atherton.
- Blalock, H. M. Jr. 1971b. Four variable causal models and partial correlations. In: Blalock, H. M. Jr. 1971a, 18-32.
- Blalock, H. M. Jr. & Blalock, A. B. (Hrsg.). 1968. Methodology in social research. New York: McGraw-Hill.
- Bock, R. D. 1975. Multivariate statistical methods in behavioral research. New York: McGraw-Hill.
- Boudon, R. 1963. Propriétés individuelles et propriétés collectives: Un problème d'analyse écologique. *Revue française de sociologie*, 4, 275-299.
- Boudon, R. 1965a. A method of linear causal analysis: Dependence analysis. *American Sociological Review*, 30, 365-374.
- Boudon, R. 1965b. Méthodes d'analyse causale. *Revue française de sociologie*, 6, 24-43.
- Boudon, R. 1968. A new look at correlation analysis. In: Blalock, H. M. & Blalock, A. B. 1968, 199-235.
- Boudon, R. 1976a. Neue Aspekte der Korrelationsanalyse. In: Hummel, H. J. & Ziegler, R. Korrelation und Kausalität, Bd. 2, 205-235.
- Boudon, R. 1976b. Anwendungen der Dependenzanalyse auf Paneldaten. In: Hummel, H. J. & Ziegler, R. Korrelation und Kausalität, Bd. 3, 421-434.
- Bortz, J. 1977. Lehrbuch der Statistik. Für Sozialwissenschaftler. Berlin: Springer.
- Brandstädter, J. & Bernitzke, F. 1976. Zur Technik der Pfadanalyse. Ein Beitrag zum Problem der nichtexperimentellen Konstruktion von Kausalmodellen. *Psychologische Beiträge*, 18, 12-34.
- Bredenkamp, J. 1969. Experiment und Feldexperiment. In: Graumann, C. F. Handbuch der Psychologie: Sozialpsychologie, Bd. 7, 1. Halbband, Theorien und Methoden. Göttingen: Hogrefe, 332-374.
- Bredenkamp, J. 1979. Das Problem der externen Validität pädagogisch-psychologischer Untersuchungen. In: Brandstädter, J., Reinert, G. & Schneewind, K. A. Pädagogische Psychologie: Probleme und Perspektiven. Stuttgart: Klett.

- Bredenkamp, J. 1980. Theorie und Planung psychologischer Experimente. Darmstadt: Steinkopff.
- Breiman, L. 1968. Probability. Reading, Mass.: Addison-Wesley.
- Brunswik, E. 1956. Perception and the representative design of psychological experiments. Berkeley: University of California Press.
- Bunge, M. 1967. Scientific research I. The search for system. Berlin: Springer.
- Campbell, D. T. 1957. Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, 54, 297-312.
- Campbell, D. T. & Stanley, J. C. 1963. Experimental and quasi-experimental designs for research on teaching. In: Gage, N. L. Handbook of research in teaching. Chicago: Rand-McNally, 171-246.
- Coleman, J. S. 1981. Longitudinal data analysis. New York: Basic Books.
- Cook, T. D. & Campbell, D. T. 1976. The design and conduct of quasi-experiments and true experiments in field settings. In: Dunette, M. D. Handbook of industrial and organizational psychology. Chicago: Rand-McNally, 223-326.
- Cook, T. D. & Campbell, D. T. 1979. Quasi-experimentation: Design & analysis issues for field settings. Chicago: Rand-McNally.
- Cox, D. R. 1970. The analysis of binary data. London: Methuen.
- Diekoppf, B. A. 1981. Die kausale Interpretierbarkeit psychologischer Aussagen. Frankfurt: Institut für Psychologie der JWG-Universität. Unveröffentlichtes Manuskript.
- Duncan, O. D. 1975. Introduction to structural equation models. New York: Academic Press.
- Essler, W. K. 1979. Wissenschaftstheorie IV. Erklärung und Kausalität. München: Alber.
- Eysenck, H. J. 1979. The structure and measurement of intelligence. Berlin: Springer.
- Fischer, G. H. 1974. Einführung in die Theorie psychologischer Tests. Bern: Huber.
- Fisher, F. M. 1966. The identification problem in econometrics. New York: McGraw-Hill.
- Fisher, F. M. 1978. Statistical identifiability. In: Kruskal, W. H. & Tanur, J. M. International encyclopedia of statistics. New York: The Free Press, Vol. II, 1066-1071.
- Fisher, R. A. 1925. Statistical methods for research workers. London: Oliver and Boyd.
- Fisher, R. A. 1947. The design of experiments. London: Oliver and Boyd.
- Fisher, R. A. 1956. Statistische Methoden für die Wissenschaft. London: Oliver and Boyd.
- Formann, A. K. 1980. Neuere Verfahren der Parameterschätzung in der Latent-Class-Analyse. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 1, 107-116.

- Gadenne, V. 1976. Die Gültigkeit psychologischer Untersuchungen. Stuttgart: Kohlhammer.
- Gaensslen, H. & Schubö, W. 1973. Einfache und komplexe statistische Analyse. Eine Darstellung der multivariaten Verfahren für Sozialwissenschaftler und Mediziner. München: Reinhardt.
- Gänssler, P. & Stute, W. 1977. Wahrscheinlichkeitstheorie. Berlin: Springer.
- Gibson, W. A. 1959. Three multivariate models: Factor analysis, latent structure analysis, and latent profile analysis. *Psychometrika*, 27, 73-81.
- Gibson, W. A. 1962. Latent structure and positive manifold. *British Journal of Statistical Psychology*, 15, 149-160.
- Goldberger, A. S. 1964. Econometric theory. New York: Wiley.
- Goldberger, A. S. 1973. Structural equation models: An overview. In: Goldberger, A. S. & Duncan, O. D., 1-18.
- Goldberger, A. S. & Duncan, O. D. 1973. (Hrsg.). Structural equation models in the social sciences. New York: Seminar Press.
- Goodman, L. A. 1974. Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61, 215-231.
- Goodman, L. A. 1978. Analyzing qualitative/categorical data. Log-linear models and latent-structure analysis. London: Addison-Wesley.
- Goodman, L. A. 1979. A note on the estimation of parameters in latent-structure analysis. *Psychometrika*, 44, 123-128.
- Granger, C. W. J. 1969. Investigating causal relations by econometric models and Cross-spectral methods. *Econometrica*, 37, 424-438.
- Granger, C. W. J. 1973. Causality, model building, and control: Some comments presented at the IFAC/FORS International Conference on Dynamit Modeling and Control, July 9-12 (University of Warwiek, Coventry).
- Green, B. F. 1951. A general solution for the latent class model of latent structure analysis. *Psychometrika*, 16, 151-166.
- Green, B. F. 1952. Latent structure analysis and its relation to factor analysis. *Journal of the American Statistical Association*, 47, 71-76.
- Hager, W. & Westermann, R. Planung und Auswertung von Experimenten. In: Breidenkamp, J. & Feger, H. Enzyklopädie der Psychologie. Forschungsmethoden der Psychologie. Bd. 5. Göttingen: Hogrefe (im Druck).
- Harnerle, A. 1982. Latent-Trait-Modelle. Weinheim: Beltz.
- Heider, F. 1958. The psychology of interpersonal relations. New York: Wiley.
- Heise, D. R. 1975. Causal analysis. New York: Wiley.
- Henning, H. J. & Muthig, K. 1979. Grundlagen konstruktiver Versuchsplanung: Ein Lehrbuch für Psychologen. München: Kösel.
- Herkner, W. H. 1980. (Hrsg.). Attribution - Psychologie der Kausalität. Bern: Huber.

- Herrmann, T. 1969. Lehrbuch der empirischen Persönlichkeitsforschung. Göttingen: Hogrefe.
- Hinderer, K. 1980. Grundbegriffe der Wahrscheinlichkeitstheorie. Berlin: Springer.
- Hodapp, V. 1978. Causal inference from nonexperimental research on anxiety and educational achievement. In: Krone, H. H. & Laux, L. Achievement, Stress, and Anxiety. Washington, D.C.: Hemisphere.
- Hosoya, Y. 1977. On the Granger condition for non-causality. *Econometrica*, 45, 1735-1736.
- Hummell, H. J. & Ziegler, R. 1976a. (Hrsg.). Korrelation und Kausalität. Bd. 1, 2 und 3. Stuttgart: Enke.
- Hummell, H. J. & Ziegler, R. 1976b. Zur Verwendung linearer Modelle bei der Kausalanalyse nicht-experimenteller Daten. In: Hummel, H. J. & Ziegler, R. Bd. 1, E 5-E 137.
- Jöreskog, K. G. 1969. A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34, 183-202.
- Jöreskog, K. G. 1970. A general method for analysis of covariance structures. *Biometrika*, 57, 239-251.
- Jöreskog, K. G. 1971. Statistical analysis of sets of congeneric tests. *Psychometrika*, 36, 109-133.
- Jöreskog, K. G. 1973. A general method for estimating a linear structural equation system. In: Goldberger, A. S. & Duncan, O. D., 85-112.
- Jöreskog, K. G. 1974. Analyzing psychological data by structural analysis of covariance matrices. In: Krantz, D. H., Atkinson, R. C., Luce, R. D. & Suppes, P. Contemporary developments in mathematical psychology. Vol. II. San Francisco: Freeman and Company, 1-56.
- Jöreskog, K. G. 1977. Structural equation models in the social sciences. Specification, estimation, and testing. In: Krishnaiah, P. R. Applications of statistics. Amsterdam: North-Holland, 265-287.
- Jöreskog, K. G. 1978. Structural analysis of covariance and correlation matrices. *Psychometrika*, 43, 443-477.
- Jöreskog, K. G. 1979. Statistical estimation of structural models in longitudinal developmental investigations. In: Nesselroade, J. R. & Baltes, P. B. Longitudinal research in the study of behavior and development. New York: Academic Press.
- Jöreskog, K. G. & Sörbom, D. 1976. Statistical models and methods for test-retest situations. In: de Gruijter, D. N. M. & van der Kamp, L. J. T. Advances in psychological and educational measurement. New York: Wiley, 285-325.
- Jöreskog, K. G. & Sörbom, D. 1977. Statistical models and methods for the analysis of longitudinal data. In: Aigner, D. J. & Goldberger, A. S., 235-285.
- Jöreskog, K. G. & Sörbom, D. 1979. Advances in factor analysis and structural equation models. Cambridge, Mass.: Abt Books.
- Jöreskog, K. G. & Sörbom, D. 1981. LISREL V. Analysis of linear structural relation-

- ships by maximum likelihood and least squares methods. University of Uppsala.  
Department of Statistics.
- Johnston, J. 1971. Econometric methods. Tokyo: McGraw-Hill Kogakusha.
- Jones, E. E., Kanouse, D. E., Kelley, H. H., Nisbett, R. E., Valins, S. & Weiner, B.:  
1971/72. Attribution: Perceiving the causes of behavior. Morristown, N. J.: General Learning Press.
- Kenny, D. A. 1979. Correlation and causality. New York: Wiley.
- Klein, L. R. 1978. Simultaneous equation estimation. In: Kruskal, W. H. & Tanur, J. M., 979-994.
- Kraak, B. 1966. Zum Problem der Kausalität in der Psychologie. Psychologische Beiträge, 9, 413-432.
- Krishnaiah, P. R. 1980. (Hrsg.). Handbook of statistics 1. Analysis of variance. Amsterdam: North-Holland.
- Kruskal, W. H. & Tanur, J. M. (Hrsg.). 1978. International encyclopedia of statistics. Vol. I and II. New York: The Free Press.
- Lawley, D. N. & Maxwell, A. E. 1971. Factor analysis as a statistical method. London: Butterworths.
- Lazarsfeld, P. F. 1955. Interpretation of statistical relations as a research Operation. In:  
Lazarsfeld, P. F. & Rosenberg, M. The language of social research. New York:  
The Free Press, 115-125.
- Lazarsfeld, P. F. 1959. Latent structure analysis. In: Koch, S. Psychology: A study of a  
science. Vol. III, New York: McGraw-Hill, 476--535.
- Lazarsfeld, P. F. & Henry, N. W. 1968. Latent structure analysis. Boston: Houghton  
Mifflin.
- Loeve, M. 1977. 1978. Probability theory. Vol. I und II. Berlin: Springer.
- Lohmöller, J. B. 1981. LVPLS 1.6 program manual. Latent variables path analysis with  
partial least-squares estimation. Forschungsbericht 81.04 des Fachbereichs  
Pädagogik der Hochschule der Bundeswehr München.
- Lord, F. M. & Novick, M. R. 1968. Statistical theories of mental test scores. Reading,  
Mass.: Addison-Wesley.
- Lumsden, J. 1976. Test theory. Annual Review of Psychology, 27, 251-280.
- Malinvaud, E. 1970. Statistical methods of econometrics. Amsterdam: North-Holland.
- Marsden, P. V. 1982. Linear models in social research. London: Sage.
- McDonald, R. P. 1962. A general approach to nonlinear factor analysis. Psychometrika, 27, 397-415.
- McDonald, R. P. 1967a. Nonlinear factor analysis. Psychometric Monograph, 15.  
Chicago: University of Chicago Press.
- McDonald, R. P. 1967b. Numerical methods for polynomial models in nonlinear  
factor analysis. Psychometrika, 32, 77-112.
- McGuigan, E. J. 1978. Experimental psychology. Englewood Cliffs: Prentice-Hall.

- Mill, J. S. 1885. System der deduktiven und induktiven Logik. Gesammelte Werke, Bd. II. Leipzig: Fues.
- Moosbrugger, H. 1978. Multivariate statistische Analyseverfahren. Stuttgart: Kohlhammer.
- Müller, P. H. 1975. (Hrsg.). Lexikon der Stochastik. Berlin: Akademie-Verlag.
- Namooddiri, N. K., Carter, L. F. & Blalock, H. M. Jr. 1975. Applied multivariate analysis and experimental designs. New York: McGraw-Hill.
- Pierce, D. A. & Haugh, L. D. 1977. Causality in temporal systems: Characterizations and a survey. *Journal of Econometrics*, 5, 265-293.
- Popper, K. R. 1959. The propensity interpretation of probability. *The British Journal for the Philosophy of Science*, 10, 25-42.
- Popper, K. R. 1969. Logik der Forschung. Tübingen: Mohr 1969.
- Preiser, S. 1977. Die experimentelle Methode. In: Strube, G. Die Psychologie des 20. Jahrhunderts. Band V, 102-150.
- Rao, C. R. 1973. Linear statistical inference and its applications. New York: Wiley.
- Rasch, G. 1960. Probabilistic models for some intelligence and attainment tests. Copenhagen: The Danish Institute for Educational Research.
- Rasch, D. 1976. Einführung in die mathematische Statistik. II. Anwendungen. Berlin: Deutscher Verlag der Wissenschaften.
- Rasch, D. 1978. Einführung in die mathematische Statistik. I. Wahrscheinlichkeitsrechnung und Grundlagen der mathematischen Statistik. Berlin: Deutscher Verlag der Wissenschaften.
- Rosenthal, R. 1964. The effect of the experimenter on the results of psychological research. In: Maher, B. A. Progress in experimental personality research. Vol. I. New York: Academic Press, 79-114.
- Rosenthal, R. 1966. Experimenter effects in behavioral research. New York: Appleton-Century-Crofts.
- Sarris, V. 1968. Zum Problem der Kausalität in der Psychologie: Ein Diskussionsbeitrag. *Psychologische Beiträge*, 10, 173-186.
- Schach, S. & Schäfer, T. 1978. Regressions- und Varianzanalyse. Eine Einführung. Berlin: Springer.
- Scheffe, H. 1959. The analysis of variance. New York: Wiley.
- Schulz, T., Muthig, K.-P. & Koeppler, K. 1981. Theorie, Experiment und Versuchsanplanung in der Psychologie. Stuttgart: Kohlhammer.
- Searle, S. R. 1971. Linear Models. New York: Wiley.
- Seibel, H. D. & Nygreen, G. T. 1972. Pfadanalyse. Ein statistisches Verfahren zur Untersuchung linearer Kausalmodelle. *Zeitschrift für Sozialpsychologie*, 3, 5-12.
- Selg, H. & Bauer, W. 1976. Forschungsmethoden der Psychologie. Stuttgart: Kohlhammer.

- Simon, H. A. 1952. On the definition of the causal relation. *The Journal of Philosophy*, 49, 517-528.
- Simon, H. A. 1953. Causal ordering and identifiability. In: Hood, W. C. & Koopmans, T. C., 49-74.
- Simon, H. A. 1954. Spurious correlation: A causal interpretation. *Journal of the American Statistical Association*, 49, 467-479.
- Spearman, C. 1904. General intelligence objectively determined and measured. *American Journal of Psychology*, 15, 201-293.
- Spearman, C. 1927. *The abilities of man*. London: Macmillan.
- Stegmüller, W. 1969. Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie. Bd. I. *Wissenschaftliche Erklärung und Begründung*. Berlin: Springer.
- Stegmüller, W. 1973. Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie. Bd. IV. *Personelle und statistische Wahrscheinlichkeit*. Berlin: Springer.
- Steyer, R. 1982. Ein Beitrag zur theoretischen Grundlegung experimenteller und nicht-experimenteller Kausalforschung. Frankfurt/M.: Dissertation am Fachbereich Psychologie der Johann Wolfgang Goethe-Universität.
- Süllwold, F. 1977. Intelligenzdiagnostik und Intelligenztheorie. In: Strube, G. *Die Psychologie des 20. Jahrhunderts*. Band V, 236-286.
- Suppes, P. 1970. *A probabilistic theory of causality*. Amsterdam: North-Holland.
- Tatsuoka, M. M. 1971. *Multivariate analysis: Techniques for educational and psychological research*. New York: Wiley.
- Timäus, E. 1974. Experiment und Psychologie. Zur Sozialpsychologie psychologischen Experimentierens. Göttingen: Hogrefe.
- Timm, N. H. 1975. *Multivariate analysis with applications in education and psychology*. Monterey, Cal.: Brooks/Cole.
- von Wright, G. H. 1974. *Erklären und Verstehen*. Frankfurt/M.: Fischer.
- Weiner, B. 1972. *Theories of motivation. From mechanism to cognition*. Chicago: Markham.
- Weiner, B. 1974. Achievement motivation as conceptualized by an attribution theorist. In: Weiner, B. *Achievement motivation and attribution theory*. Morristown, N. J.: General Learning Press.
- Wiley, D. E. 1973. The identification problem for structural equation models with unmeasured variables. In: Goldberger, A. S. & Duncan, O. D., 69-83.
- Wilkening, K. & Wilkening, F. 1979. *Versuchsplanung*. Tübingen: Versuch für das Fernstudium im Medienverbund. Kapitel 3.
- Winer, B. J. 1971. *Statistical principles in experimental design*. New York: McGraw-Hill.
- Wold, H. 1954. Causality and econometrics. *Econometrica*, 22, 162-177.

- Wold, H. 1956. Causal inference from observational **data**. A review of ends and means. *Journal of the Royal Statistical Society (A)*, 119, 28-61.
- Wold, H. 1964. (Hrsg.). Model building in the human sciences. Monaco: Union europeenne d'éditions.
- Wold, H. 1969. Mergers of economics and philosophy of science. *Synthese*, 20, 427-482.
- Wold, H. 1981. (Hrsg.). The fix-point approach to interdependent systems. Amsterdam: North-Holland.
- Wright, S. 1918. On the nature of size factors. *Genetics*, 3, 367-374.
- Wright, S. 1921. Correlation and Causation. *Journal of Agricultural Research*, 20, 557-585.
- Wright, S. 1923. The theory of path coefficients - A reply to Niles' criticism. *Genetics*, 8, 239-255.
- Wright, S. 1934. The method of path coefficients. *Annals of Mathematical Statistics*, 5, 161-215.
- Wright, S. 1960a. Path coefficients and path regressions: alternative or complementary concepts? *Biometrics*, 16, 189-202.
- Wright, S. 1960b. The treatment of reciprocal interaction, with or without lag, in path analysis. *Biometrics*, 16, 423-445.
- Zimmermann, E. 1972. Das Experiment in den Sozialwissenschaften. Stuttgart: Teubner.

Der vorliegende Beitrag ist eine revidierte Version meiner mit ‚Steyer (1982)‘ zitierten Dissertation. Den Gutachtern möchte ich für wertvolle Diskussionen und kritische Anmerkungen danken: Prof. Dr. W. Bauer, Prof. Dr. H. Dinges, Prof. Dr. H. Moosbrugger und Prof. N. Wermuth, PhD.

### 3. Kapitel

## Uni- und multivariate Varianzanalyse mit festen Parametern

*Helfried Moosbrugger und Rolf Steyer*

### 1. Einführung und Überblick

Eine der am häufigsten auftretenden Fragen in der bio- und sozialwissenschaftlichen Forschung ist die, ob Mittelwerteunterschiede zwischen zwei oder mehr als zwei Gruppen bezüglich eines bestimmten „gemessenen“ Merkmals bestehen. Diese Frage kann durch einfaches Vergleichen der Mittelwerte in den Gruppen nicht zufriedenstellend entschieden werden, wenn nur eine *Stichprobe* von Gruppenmitgliedern untersucht wurde, aber dennoch eine Aussage für die *Population*, d.h. für alle Gruppenmitglieder getroffen werden soll.

Die erste exakte Lösung für den Fall zweier Gruppen wurde 1908 von Gosset unter dem Pseudonym „Student“ (vgl. z.B. Pearson & Wishart, 1943) veröffentlicht und ist unter der Bezeichnung *t-Test* auch heute noch eines der am häufigsten verwendeten inferenzstatistischen Verfahren.

Für den Fall mehrerer Gruppen hat Fisher (1925, 1935) die „grundlegenden Prinzipien der Varianzanalyse und ihre wichtigsten Techniken ausgearbeitet und publiziert“ (Stanley, 1978, S. 541, Übersetzung durch die Autoren). Schließlich wurde von Roy (z.B. 1939, 1946, 1957) die *multivariate Varianzanalyse* entwickelt (vgl. Stanley, 1978, S. 553), mit der man Mittelwerteunterschiede mehrerer Gruppen hinsichtlich *mehrerer* „gemessener“ Merkmale untersuchen kann.

Seit Cohens (1968) Artikel über „Multiple Regression als ein allgemeines System zur Datenanalyse“ (Übersetzung durch die Autoren) ist in den Fachzeitschriften der Psychologie eine Vielzahl von Arbeiten erschienen, die auf die allgemeine Theorie linearer Modelle eingehen, in welche die Varianzanalyse, wie z.B. auch die Regressionsanalyse, eingebettet werden kann (siehe z.B. Moosbrugger, 1978, oder Steyer, 1979, und die dort angegebene Literatur).

Die Theorie linearer Modelle war aber bereits vorher in der mathematisch-statistischen Literatur bekannt und weit entwickelt (siehe z.B. Anderson, 1958, Graybill, 1961, oder Scheffé, 1959).

Varianz- und Regressionsanalysen als Spezialfälle ein- und derselben Theorie zu behandeln wird dadurch möglich, daß sich Mittelwerteunterschiede einer Variablen in z.B. zwei Gruppen auch als Zusammenhang der betreffenden Variablen mit einer zweistufigen Kodiervariablen, welche die Gruppenzugehörigkeit anzeigt, darstellen lassen. Zusammenhänge zwischen zwei Variablen aber werden gewöhnlich mit Methoden der Regressions- und Korrelationsanalyse untersucht.

In dem vorliegenden Beitrag stellen wir die u.E. wichtigsten Begriffe der Theorie multivariater linearer Modelle mit festen Parametern dar und zeigen exemplarisch, wie auf dieser Basis verschiedene varianzanalytische Fragestellungen zu bearbeiten sind. Gegenüber der herkömmlichen Darstellung hat dies vor allem didaktische und konzeptuelle Vorteile; didaktische, weil nur wenige Grundbegriffe vermittelt werden müssen, und konzeptuelle, weil der Ansatz sehr allgemein ist.

Im zweiten Abschnitt stellen wir die *Grundbegriffe* der hier behandelten Theorie dar. Im dritten Abschnitt führen wir die *multivariate allgemeine lineare Hypothese* ein und erläutern deren Anwendung an einer Reihe verschiedener *Designs*. Dabei legen wir das „Zellenmittelwertemodell“ zugrunde, bei dem die Parametermatrix eines linearen Modells die Zellenmittelwerte der abhängigen Variablen als Komponenten enthält.

Im vierten Abschnitt behandeln wir dann Fragen der *Parameterschätzung* und im fünften schließlich solche der *Hypothesenbewertung*, wobei wir beide Themenkomplexe für den allgemeinen Fall abhandeln, bei dem die Matrix der unabhängigen Variablen nicht vollen Spaltenrang haben muß.

In allen Abschnitten setzen wir einige Grundkenntnisse der Matrixalgebra voraus, in welche z.B. Moosbrugger (1978), Searle (1966) oder Van de Geer (1971) einführen.

## **2. Multivariate lineare Modelle mit festen Parametern**

### **2.1 Einleitung**

In diesem zweiten Abschnitt wird diejenige Klasse von linearen statistischen Modellen formal charakterisiert, mit denen wir uns im weiteren Verlauf dieses Kapitels beschäftigen. Dabei werden wir zunächst in die grundlegenden Ideen

einführen und dann die Verbindung dieser Ideen mit Beobachtungseinheiten (z.B. Personen) und Stichproben herstellen.

Von vielen Autoren wird die hier dargestellte Theorie unter der Bezeichnung „Allgemeines (multivariates) lineares Modell“ abgehandelt (siehe z.B. Finn, 1974, Graybill, 1976, Horton, 1978, Mendenhall, 1968, Moosbrugger, 1978, Steyer, 1979), eine Bezeichnung, die Verwirrung stiften könnte, da sie die Annahme nahelegt, es handele sich hier um die allgemeinste Form linearer Modelle. Dies ist aber nicht der Fall, was man schon daran erkennen kann, daß etwa die „random“ und „mixed effects“ Modelle (vgl. z.B. Searle, 1971, S. 376ff.) nicht unter diese Kategorie fallen, genausowenig wie lineare Modelle mit stochastischen Prädiktoren (vgl. z.B. Graybill, 1976, Kap. 10).

## 2.2 Die grundlegenden Modellvorstellungen

In diesem Beitrag betrachten wir ausschließlich Modelle, bei denen für die stochastischen Vektoren  $y = (y_1, y_2, \dots, y_p)$  von „abhängigen“ Variablen und  $x = (x_1, x_2, \dots, x_Q)$  von „unabhängigen“ Variablen die Gleichung

$$(2.2.1) \quad E(y | x) = x\beta$$

( $\beta$  ist ein großes griechisches Beta) gilt, welche die  $P$  einzelnen Gleichungen

$$(2.2.2) \quad E(y_p | x_1, x_2, \dots, x_Q) = x_p\beta_p = \beta_{p1}x_1 + \beta_{p2}x_2 + \dots + \beta_{pQ}x_Q$$

enthält. Demnach ist die bedingte Erwartung<sup>1)</sup> einer abhängigen Variablen  $y_p$ ,  $p \in \{1, 2, \dots, P\}$ , unter  $x_q$ ,  $q \in \{1, 2, \dots, Q\}$ , eine mit den festen Parametern  $\beta_{pq}$  gewichtete Summe der unabhängigen Variablen  $x_q$ . Die Koeffizienten  $\beta_{pq}$ , die auch partielle Regressionskoeffizienten genannt werden (vgl. z.B. Gaensslen & Schubö, 1973, S. 92) geben an, wie stark erwartungsgemäß zwei Werte von  $y_p$  differieren, wenn sich die zugehörigen Werte von  $x_q$  um den Betrag Eins unterscheiden, falls die Werte aller anderen unabhängigen Variablen  $x_{q^*}$ ,  $q^* \neq q$ , gleich bleiben.

In der Varianzanalyse handelt es sich bei jedem  $y_p$ ,  $p \in \{1, 2, \dots, P\}$ , um eine quantitative Variable und bei den  $x_q$ ,  $q \in \{1, 2, \dots, Q\}$ , um qualitative Variablen, welche Gruppenzugehörigkeiten durch Zahlen wie z.B. 0 oder 1 anzeigen.

<sup>1)</sup> Man beachte, daß es sich bei der bedingten Erwartung um eine stochastische Variable handelt, im Unterschied zum bedingten Erwartungswert. Zur Definition dieser Begriffe der Wahrscheinlichkeitstheorie verweisen wir auf Bauer (1974) oder Müller (1975). Gleichungen für bedingte Erwartungen gelten nur mit Wahrscheinlichkeit 1, worauf wir im folgenden nicht weiter hinweisen.

Wir erläutern nun Gleichung 2.2.1 für einige bekannte statistische Modelle. Das einfachste aller denkbaren Modelle (mit P=1 und Q=1) ist

$$(2.2.3) \quad E(y | x) = \beta \cdot 1 = \beta,$$

welches auch in der Form

$$(2.2.4) \quad E(y) = \mu$$

geschrieben werden kann. Damit wird die Annahme formuliert, daß eine stochastische Variable  $y$  einen Erwartungswert hat, nämlich  $\mu$  bzw.  $\beta$ . Die stochastische Variable  $x$  ist hier zu der Konstanten Eins „degeneriert“, weswegen in diesem Fall auch  $E(y | x) = E(y)$  gilt. In einer Stichprobensituation ist hiermit oft die bekannte Fragestellung verbunden, ob ein Stichprobenmittelwert „signifikant“ von  $\mu$  abweicht (vgl. z.B. Bortz, 1977, S. 156-160).

Ein ebenso bekanntes Modell mit  $P=1$  und  $Q=2$  liegt vor, wenn der stochastische Vektor  $x$  in Gleichung 2.2.5 nur die beiden Ausprägungen

$$x_1 = [1 \ 0] \text{ und } x_2 = [0 \ 1]$$

annehmen kann:

$$(2.2.5) \quad E(y | x) = x\beta = \beta_1 x_1 + \beta_2 x_2.$$

Es ist dann mit den beiden Gleichungen

$$(2.2.6) \quad E(y | x_{11}=1, x_{21}=0) = \beta_1 \cdot 1 + \beta_2 \cdot 0 = \beta_1$$

und

$$(2.2.7) \quad E(y | x_{12}=0, x_{22}=1) = \beta_1 \cdot 0 + \beta_2 \cdot 1 = \beta_2$$

gleichbedeutend.

Dieses Modell kann man auch mit den Parametern  $\mu_1$  statt  $\beta_1$  und  $\mu_2$  statt  $\beta_2$  schreiben, wobei  $\mu_1$  als Erwartungswert der abhängigen Variablen  $y$  in einer ersten und  $\mu_2$  als deren Erwartungswert in einer zweiten Gruppe zu interpretieren ist. In einer Stichprobensituation ist mit diesem Modell oft die Fragestellung verbunden, ob zwei Stichprobenmittelwerte „signifikant“ voneinander abweichen (vgl. z.B. Bortz, 1977, S. 160-164). Für  $Q=3$  lautet die Modellgleichung

$$(2.2.8) \quad E(y | x) = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

Im Fall dreier Gruppen hat der stochastische Vektor  $x$  nur die drei Ausprägungen

$$\begin{aligned}\mathbf{x}_1 &= [1 \ 0 \ 0], \\ \mathbf{x}_2 &= [0 \ 1 \ 0] \\ \text{und } \mathbf{x}_3 &= [0 \ 0 \ 1]\end{aligned}$$

Gleichung (2.2.7) ist dann mit den drei Gleichungen

$$E(y | x_{11}=1, x_{21}=0, x_{31}=0) = \beta_1 \cdot 1 + \beta_2 \cdot 0 + \beta_3 \cdot 0 = \beta_1$$

$$E(y | x_{12}=0, x_{22}=1, x_{32}=0) = \beta_1 \cdot 0 + \beta_2 \cdot 1 + \beta_3 \cdot 0 = \beta_2$$

$$E(y | x_{13}=0, x_{23}=0, x_{33}=1) = \beta_1 \cdot 0 + \beta_2 \cdot 0 + \beta_3 \cdot 1 = \beta_3$$

oder

$$(2.2.9) \quad E(y | \mathbf{x}=\mathbf{x}_q) = \beta_q, \quad q \in \{1,2,3\},$$

gleichbedeutend.

Eine verbreitetere Schreibweise dafür ist jedoch

$$(2.2.10) \quad E(y | \mathbf{x}=\mathbf{x}_q) = \mu_q = \mu + (\mu_q - \mu) = \mu + \alpha_q, \quad q \in \{1,2,3\},$$

wobei  $\alpha_q := \mu_q - \mu$ . Mit diesem Modell ist in Stichprobensituationen oft die *varianzanalytische Fragestellung* verbunden, ob zwischen den Q=3 Gruppen „signifikante“ Mittelwerteunterschiede bestehen.

Nachdem wir mit diesen einfachen Beispielen die Gleichung 2.2.1 erläutert haben, kehren wir zur Darstellung der allgemeinen Theorie zurück.

Aus der Gleichung 2.2.1 lassen sich eine Reihe von Eigenschaften eines linearen Modells ableiten. Betrachten wir z.B. den *Residualvektor* vom Typ 1xP

$$(2.2.11) \quad \boldsymbol{\epsilon} := \mathbf{y} - E(\mathbf{y} | \mathbf{x})$$

( $\boldsymbol{\epsilon}$  ist ein kleines griechisches Epsilon), so gelten dafür nach den Regeln 4, 5 und 10 des Anhangs C von Moosbrugger (1982), in diesem Band,

$$(2.2.12) \quad E(\boldsymbol{\epsilon} | \mathbf{x}) = \mathbf{0},$$

nach Regel 1 jenes Anhangs

$$(2.2.13) \quad E(\boldsymbol{\epsilon}) = \mathbf{0}$$

und nach Regel 11 jenes Anhangs

$$(2.2.14) \quad E(\mathbf{x}' \boldsymbol{\epsilon}) = C(\mathbf{x}, \boldsymbol{\epsilon}) = \mathbf{O}$$

wobei  $\mathbf{x}'$  ein Spaltenvektor vom Typ Q x 1 ist.

Nach Gleichung 2.2.12 ist die bedingte Erwartung von  $\boldsymbol{\varepsilon}$  unter  $x = (x_1, x_2, \dots, x_Q)$  gleich Null, wie auch der unbedingte Erwartungswert von  $\boldsymbol{\varepsilon}$  (Gleichung 2.2.13). Wenn die bedingte Erwartung von  $\boldsymbol{\varepsilon}$  gegeben  $x_1, x_2, \dots, x_Q$  und damit auch die bedingten Erwartungswerte für alle Ausprägungskombinationen der  $x_q$  gleich Null sind, dann kann auch keine Residualvariable  $\varepsilon_p$  mit einem  $x_q$  kovariieren (Gleichung 2.2.14) oder korrelieren.

Aus der Gleichung 2.2.1 folgt, wenn  $E(x'x)$  nonsingulär ist, die *Identifikationsgleichung* für die Parametermatrix B (vgl. Moosbrugger, 1982, in diesem Band, Kap. 2.2.3).

$$(2.2.15) \quad \mathbf{B} = E(x'x)^{-1} E(x'y).$$

Von praktischem Interesse sind neben den Parametern in der Matrix B, die u.a. zur *Prädiktion* dienen können, auch die durch die  $x_1, x_2, \dots, x_Q$ , also durch  $x$  bestimmte Varianz  $V[E(y_p | x)]$  einer „abhängigen“ Variablen  $y_p$ , für die nach Gleichung 2.2.2

$$(2.2.16) \quad V[E(y_p | x)] = C(x\beta_p, x\beta_p) =$$

$$(2.2.17) \quad = \beta'_p C(x, x) \beta_p =$$

$$(2.2.18) \quad = V(y_p) - V(\varepsilon_p)$$

gilt (siehe Gleichung B.5 und Regel 2 des Anhangs B von Moosbrugger, 1982, in diesem Band). Sie ist ein Kennwert für die *Güte der Prädiktion* oder für die *Stärke des Zusammenhangs* zwischen  $y_p$  und  $x$ , ebenso wie der durch  $x$  bestimmte Varianzanteil von  $y_p$ , der *multiple Determinationskoeffizient*

$$(2.2.19) \quad R^2_{y_p|x} := \frac{V[E(y_p | x)]}{V(y_p)} = \frac{V(y_p) - V(\varepsilon_p)}{V(y_p)} .$$

Dieser ist von großem praktischen Interesse, weil er ein *normiertes Maß* für die *praktische Bedeutsamkeit* (vgl. z.B. Bredenkamp, 1970) eines statistischen Zusammenhangs zwischen  $y_p$  und  $x$  ist, insofern, als  $R^2_{y_p|x}$  100% den durch  $x$  bestimmten Prozentanteil der Varianz von  $y_p$  angibt.

Die positive Wurzel von  $R^2_{y_p|x}$  heißt *multipler Korrelationskoeffizient*, der ebenso wie der multiple Determinationskoeffizient Werte zwischen Null und Eins annehmen kann. Der Wert Eins steht dabei für einen perfekten Zusammenhang, der eine fehlerfreie Vorhersage von  $y_p$  aufgrund der Werte der  $x_q$  ermöglicht, und  $R^2_{y_p|x}$  gleich Null besagt, daß die Kenntnis der Werte der  $x_q$  keine bessere Vorhersage von  $y_p$  ermöglicht als deren Erwartungswert  $E(y_p)$ . Dies ist immer dann der Fall, wenn alle Parameter  $\beta_{pq}$ ,  $q \in \{1, 2, \dots, Q\}$ , gleich Null sind, und kein Unterschied zwischen der bedingten Erwartung  $E(y_p | x)$  und dem (unbedingten) Erwartungswert  $E(y_p)$  besteht.

Skalare Kennwerte der praktischen Bedeutsamkeit für den Fall *mehrerer* abhängiger Variablen ( $P \geq 2$ ) sind weniger eindeutig anzugeben, da zunächst nicht nur ein einziger Kennwert für die Streubreite einer abhängigen Variablen  $y_p$ , nämlich die Varianz  $V(y_p)$ , vorliegt, sondern eine symmetrische  $P \times P$ -Kovarianzmatrix der Komponenten von  $y$ . Analog zum univariaten Fall kann man zunächst jedoch *die durch  $x$  bestimmte Kovarianzmatrix von  $y$*

$$(2.2.20) \quad C[E(y|x), E(y|x)] = C[xB, xB] = B'C(x,x) B =$$

$$(2.2.21) \quad = C(y,y) - C(\epsilon,\epsilon)$$

angeben (siehe Gleichung B.5 und Regel 2 des Anhangs B von Moosbrugger, 1982, in diesem Band). Ebenso wie im univariaten Fall können wir diese Kovarianzmatrizen auf die Kovarianzmatrix  $C(y,y)$  der abhängigen Variablen normieren, indem wir die Gleichung 2.2.21 mit  $C(y,y)^{-1}$  nachmultiplizieren, wodurch wir bei Nonsingularität von  $C(y,y)$  die zu 2.2.19 analoge Gleichung

$$(2.2.22) \quad C[E(y|x), E(y|x)] C(y,y)^{-1} = [C(y,y) - C(\epsilon,\epsilon)] C(y,y)^{-1}$$

erhalten. Ausmultiplizieren und Umstellen ergibt

$$(2.2.23) \quad C[E(y|x), E(y|x)] C(y,y)^{-1} + C(\epsilon,\epsilon) C(y,y)^{-1} = I,$$

woraus man ersieht, daß sich der durch  $x$  bestimmte Anteil der Kovarianzmatrix (der linke Summand) und deren unbestimmter Anteil (der rechte Summand) zur Einheitsmatrix  $I$  aufaddieren. Ausgehend von diesen Matrizen wurden eine Reihe von multivariaten Verallgemeinerungen des Determinationskoeffizienten vorgeschlagen, die z.B. bei Shaffer und Gillo (1974) diskutiert werden. Am plausibelsten erscheint uns jedoch der Vorschlag von Cramer und Nicewander (1979), die den Schätzer für den Kennwert

$$(2.2.24) \quad R_{y|x}^2 := \frac{\text{Spur}[C[E(y|x), E(y|x)] C(y,y)^{-1}]}{P}$$

als *multivariaten Determinationskoeffizienten* favorisieren. Dabei ist

$$(2.2.25) \quad P = \text{Spur}[C(y,y) C(y,y)^{-1}] = \text{Spur } I$$

die Anzahl der Komponenten von  $y$  und zugleich der Rang von  $C(y,y)$ . Dieser Kennwert ist invariant gegenüber linearen Skalentransformationen von  $y$ , liegt zwischen Null und Eins, reduziert sich im univariaten Fall auf den Determinationskoeffizienten  $R_{y|x}^2$  (siehe Gleichung 2.2.19) und addiert sich mit

$$\frac{\text{Spur}[C(\epsilon,\epsilon) C(y,y)^{-1}]}{P}$$

zu Eins auf. Außerdem besteht ein sehr enger Zusammenhang dieses multivariaten Determinationskoeffizienten zum Pillai-Bartlett Spur-Kriterium, wel-

ches wir neben anderen multivariaten Testkriterien im Abschnitt 5 behandeln werden.

## 2.3 Stichprobenmodelle

In Abschnitt 2.2 haben wir die lineare Modellgleichung

$$(2.3.1) \quad E(\mathbf{y} | \mathbf{x}) = \mathbf{x}\mathbf{B}$$

formuliert. Es stellt sich nun die Frage, wie man die in der Regel unbekannten Parameter in der Matrix B schätzen kann. In der psychologischen Forschung geben die Werte der stochastischen Variablen in y und x Merkmale von Beobachtungseinheiten, z.B. Versuchspersonen, wieder. Bei den abhängigen Variablen handelt es sich um Messungen einer ( $P=1$ ) oder mehrerer ( $P>1$ ) Eigenschaften einer Beobachtungseinheit, und bei den unabhängigen Variablen z.B. um Indikatoren der Zugehörigkeit zu einer Gruppe (z. B. Experimental- oder Kontrollgruppe).

Um ein Verfahren zur Schätzung der Parametermatrix B entwickeln zu können, müssen wir unsere Vorstellungen über die uns vorliegenden N Beobachtungseinheiten präzisieren. Die Beobachtungseinheiten können entweder verschiedene Vpn sein oder Person-Zeit-Kombinationen, nämlich Personen zu verschiedenen Meßzeitpunkten. Zunächst legen wir fest, daß die Gleichung 2.3.1 für jede der N Beobachtungseinheiten gelten soll:

$$(2.3.2) \quad E(\mathbf{y}_n | \mathbf{x}_n) = \mathbf{x}_n\mathbf{B}, \quad \text{für alle } n \in \{1, 2, \dots, N\}.$$

Für varianzanalytische Fragestellungen liegt typischerweise mit jeder Beobachtungseinheit auch deren Gruppenzugehörigkeit fest und somit auch die Realisierung x, des Vektors der unabhängigen Variablen  $\mathbf{x}_n$ .<sup>2)</sup> Bei den abhängigen Variablen in  $\mathbf{y}_n$  hingegen liegen in der Varianzanalyse bei gegebener Beobachtungseinheit n die Realisierungen  $y_n$  des stochastischen Vektors  $\mathbf{y}_n$  noch nicht fest, d.h. auch bei gegebener Beobachtungseinheit ließe sich prinzipiell eine Variation der betreffenden Variablen feststellen.

Da bei gegebener Beobachtungseinheit n der stochastische Vektor  $\mathbf{x}_n$  mit seiner Realisierung  $x_n$  identisch ist, gilt außer  $E(\mathbf{y}_n | \mathbf{x}_n) = \mathbf{x}_n\mathbf{B}$  auch

$$(2.3.3) \quad E(\mathbf{y}_n | \mathbf{x}_n) = E(\mathbf{y}_n | \mathbf{x}_n = \mathbf{x}_n) = E(\mathbf{y}_n)$$

und

$$(2.3.4) \quad E(\mathbf{y}_n) = E[E(\mathbf{y}_n | \mathbf{x}_n)] = E(\mathbf{x}_n\mathbf{B}) = \mathbf{x}_n\mathbf{B}, \quad n \in \{1, 2, \dots, N\}$$

---

<sup>2)</sup> Diese Eigenschaft unterscheidet die hier behandelten Modelle von solchen mit „stochastischen Regressoren“, über welche z.B. Graybill, 1976, S. 373ff., informiert.

(siehe Regel 1 des Anhangs C und Regel 1 des Anhangs A von Moosbrugger, 1982, in diesem Band).

Aus Gleichung 2.3.3 und 2.3.4 ist ersichtlich, daß das Stichprobenmodell einfach wie folgt geschrieben werden kann:

$$(2.3.5) \quad E(y_n) = x_n B, \quad \text{für alle } n \in \{1, 2, \dots, N\}.$$

Es ist nützlich diese N-Gleichungen mit einer Matrizengleichung zu notieren, wobei die folgenden Matrizen verwendet werden:

Y bestehend aus  $N \times P$  stochastischen Variablen,

$$(2.3.6) \quad Y = \begin{bmatrix} y_{11}, \dots, y_{1p}, \dots, y_{1P} \\ \vdots & \vdots & \vdots \\ y_{n1}, \dots, y_{np}, \dots, y_{nP} \\ \vdots & \vdots & \vdots \\ y_{N1}, \dots, y_{Np}, \dots, y_{NP} \end{bmatrix},$$

X bestehend aus  $N \times Q$  festen reellen Zahlen

$$(2.3.7) \quad X = \begin{bmatrix} x_{11}, \dots, x_{1q}, \dots, x_{1Q} \\ \vdots & \vdots & \vdots \\ x_{n1}, \dots, x_{nq}, \dots, x_{nQ} \\ \vdots & \vdots & \vdots \\ x_{N1}, \dots, x_{Nq}, \dots, x_{NQ} \end{bmatrix}$$

und

B bestehend aus  $Q \times P$  festen reellen Zahlen

$$(2.3.8) \quad B = \begin{bmatrix} \beta_{11}, \dots, \beta_{1p}, \dots, \beta_{1P} \\ \vdots & \vdots & \vdots \\ \beta_{q1}, \dots, \beta_{qp}, \dots, \beta_{qP} \\ \vdots & \vdots & \vdots \\ \beta_{Q1}, \dots, \beta_{Qp}, \dots, \beta_{QP} \end{bmatrix}.$$

Für alle N Beobachtungseinheiten zusammen lautet das Stichprobenmodell dann in Matrzenschreibweise

$$(2.3.9) \quad E(\mathbf{Y}) = \mathbf{X} \mathbf{B}.$$

Die Annahme der Gleichheit der Parametermatrix B für alle Beobachtungseinheiten in Gleichung 2.3.9 wird noch durch eine Annahme über die Kovarianzmatrix der Residualvektoren

$$(2.3.10) \quad \boldsymbol{\epsilon}_n := \mathbf{y}_n - \mathbf{x}_n \mathbf{B}$$

ergänzt, nämlich durch

$$(2.3.11) \quad C(\boldsymbol{\epsilon}_n, \boldsymbol{\epsilon}_{n^*}) = \begin{cases} \boldsymbol{\Sigma}, & \text{für } n = n^* \\ \mathbf{O}, & \text{für } n \neq n^* \end{cases} \quad \text{für alle } n, n^* \in \{1, 2, \dots, N\}.$$

Dies bedeutet, daß die PxP-Kovarianzmatrix  $\boldsymbol{\Sigma}$  der Residuen der abhängigen Variablen für jede der N Beobachtungseinheiten gleich ist und daß die PxP Kovarianzmatrizen der Residuen der abhängigen Variablen verschiedener Beobachtungseinheiten als Nullmatrizen angenommen werden.<sup>3)</sup>

Ob es zweckmäßig ist, mit dem Index n eine Person oder eine Person-Zeit-Kombination zu bezeichnen, entscheiden wir in einem Anwendungsfall danach, ob dann Gleichung 2.3.11 angenommen werden kann oder nicht. Man kann annehmen, daß die Kovarianzmatrix  $C(\boldsymbol{\epsilon}_n, \boldsymbol{\epsilon}_{n^*})$  für zwei verschiedene Beobachtungseinheiten  $n \neq n^*$  z.B. dann gleich der Nullmatrix ist, wenn  $y_n$  und  $y_{n^*}$ , die abhängigen Variablen von zwei verschiedenen Personen enthält und eine gegenseitige Beeinflussung, wie z.B. Abschreiben bei einem Test, ausgeschlossen ist. Wenn  $y_n$  und  $y_{n^*}$  Beobachtungen der gleichen Person zu verschiedenen Zeitpunkten enthält, kann diese Annahme u. U. auch als erfüllt angesehen werden. Bei den herkömmlichen Designs mit wiederholten Messungen (vgl. z.B. Hays, 1973, S. 568ff.) beispielsweise, macht man diese Annahme.

Sind wir bereit, die Annahme der Gleichung 2.3.11 zu akzeptieren, wenn n eine Person-Zeit-Kombination indiziert, dann schreiben wir die Zeilenvektoren der abhängigen Variablen für die verschiedenen Zeitpunkte der Messungen einer Person untereinander, d.h. der Index n weist dann nicht auf eine Person, sondern auf die Person zum Zeitpunkt t hin.

Können wir aus theoretischen Überlegungen oder auf Grund eines statistischen Tests die Gleichung 2.3.11 für Person-Zeit-Kombinationen n nicht akzeptieren, so fassen wir die an den Personen wiederholten Messungen als Realisationen von (formal gesehen) weiteren abhängigen Variablen auf. Hätten

---

<sup>3)</sup> Im univariaten Fall ( $P=1$ ) reduziert sich die Matrix  $\boldsymbol{\Sigma}$  zu dem Skalar  $\sigma^2 = \sigma_{11}$ .

wir zunächst 2 abhängige Variablen dreimal gemessen, so hätten wir nach der Umstrukturierung  $2 \times 3 = 6$  abhängige Variablen vorliegen. Gemäß Gleichung 2.3.11 dürfen die Residuen dieser 6 abhängigen Variablen miteinander korrelieren, wobei  $\Sigma$  deren Kovarianzmatrix ist.

Die Gleichungen 2.3.9 bis 2.3.11 werden in der Literatur häufig als „*multivariate Gauss-Markoff setup*“ und Gleichung 2.3.9 als „*multivariates allgemeines lineares Modell*“ bezeichnet. In diesem Artikel werden wir uns ausschließlich mit Modellen befassen, welche diese Gleichungen erfüllen. Für einige andere Modelle geben wir Literaturhinweise.

## 2.4 Zusammenfassende Bemerkungen

In diesem zweiten Abschnitt haben wir die grundlegenden Vorstellungen der hier behandelten Klasse von linearen statistischen Modellen dargestellt. Dabei handelt es sich um die lineare Modellgleichung, die Identifikationsgleichung, die Begriffe bestimmte Varianz und Determinationskoeffizient sowie um deren Verallgemeinerung im multivariaten Fall, bei dem mehrere abhängige Variablen vorliegen.

Auf der Stichprobenebene sind die wesentlichen Annahmen, daß mit einer Beobachtungseinheit auch die Realisation  $x$  des stochastischen Vektors  $x$  der unabhängigen Variablen festliegt („nicht-stochastische Regressoren“), daß die Parametermatrix  $B$  nicht-stochastische Koeffizienten enthält („fixed-effects“), die für alle Beobachtungseinheiten gleichermaßen gelten, daß die Kovarianzmatrix der Residuen der  $P$  abhängigen Variablen bei gegebener Beobachtungseinheit  $n$  für alle  $N$  Beobachtungseinheiten gleich ist (siehe Gleichung 2.3.11), und daß diese Residuen für unterschiedliche Beobachtungseinheiten unkorreliert sind.

*Weiterführende Literatur:* Die Literatur über lineare Modelle oder das „Allgemeine lineare Modell“ ist im letzten Jahrzehnt fast unübersehbar geworden. Zunächst seien ohne Anspruch auf Vollständigkeit einige Monographien genannt, die zur Einführung in die hier behandelte stochastische Theorie dienen können: Bock (1975), Cohen und Cohen (1975), Edwards (1976), Ezekiel und Fox (1959), Finn (1974), Fraser (1979), Gaensslen und Schubö (1976), Graybill (1976), Horton (1978), Krafft (1978), Lindeman et al. (1980), Mendenhall (1968), Moosbrugger (1978), Namboodiri et al. (1975), Overall und Klett (1972), Rao (1973), Schach und Schäfer (1978), Searle (1971), Seber (1980), Tatsuoka (1971), Timm (1975), Ward und Jennings (1973) sowie Winer (1971).

Zu Einführung in lineare Modelle mit stochastischen Parametern („random effects“) seien Ahrens und Läuter (1981), Hild (1977) und Searle (1971, Kap. 9-11), genannt, Modelle mit stochastischen Regressoren behandeln u.a. Graybill (1976) oder Johnston (1971).

### 3. Hypothesenformulierung in verschiedenen Designs

#### 3.1 Einleitung

Die Schätzung von Parametern, Berechnung von Teststatistiken, Überschreitungswahrscheinlichkeiten etc., zur hier dargestellten allgemeinen Theorie sind heute, sofern Rechenanlagen verfügbar sind, weitgehend automatisiert durchführbar und für den Psychologen daher nicht von *primärem* Interesse. Dieser muß lediglich entscheiden, ob die dabei vorausgesetzten Annahmen erfüllt sind. Unabdingbar wichtig ist für jeden Psychologen jedoch das Wissen, welche Hypothesen er in welcher Situation untersuchen soll. Im folgenden werden wir daher zunächst auf die Hypothesenformulierung für verschiedene Designs eingehen, womit zugleich ein Eindruck vermittelt werden soll, wie weit gespannt der Anwendungsbereich der hier dargestellten Theorie ist, aber auch, wo dessen Grenzen liegen. Die Formulierung und Formalisierung von Hypothesen ist außerdem jener Bereich, der am wenigstens automatisiert werden kann und sollte.

Bei der Behandlung der Designs werden wir vom *Zellenmittelwertemodell* ausgehen, bei dem die Parameter in der Matrix B als bedingte Erwartungswerte der abhängigen Variablen in den Zellen des Designs interpretiert werden können. Die Besonderheiten dieses Modells werden wir im Abschnitt 3.2 behandeln. Im Abschnitt 3.3 wird die allgemeine multivariate lineare Hypothese eingeführt, und ihre Anwendung in den Abschnitten 3.4 bei 3.9 an verschiedenen Designs erläutert. Dabei gehen wir auf kreuzfaktorielle Designs, solche mit hierarchischen Faktoren und auf lateinische Quadrate ein. Bei allen Designs setzen wir gleiche Zellenfrequenzen voraus. Für nonorthogonale Designs sei auf die entsprechende Spezialliteratur, z.B. Appelbaum und Cramer (1974), Herr und Gaebelein (1978), Overall und Spiegel (1973a,b), Overall, Spiegel und Cohen (1975), Rawlings (1972, 1973) oder Steyer (1979), hingewiesen.

#### 3.2 Das Zellenmittelwertemodell

Bei allen im folgenden behandelten Designs werden wir das „Zellenmittelwertemodell“ zugrunde legen, in welchem die Parameter  $\beta_{qp}$  in der Matrix B nichts anderes als die Erwartungswerte  $p_{qp}$  der Variablen  $y_p$  in der q-ten Zelle eines Designs sind.<sup>4)</sup>

<sup>4)</sup> Timm und Carlson (1975) benutzen für diesen Ansatz die Bezeichnung „Modell vollen Ranges“, die wir aber für ungeeignet halten, da noch viele andere Modelle existieren, bei denen die Matrix X vollen Rang hat, bei denen also die Anzahl der Spalten von X gleich dem Rang von X ist, ohne die genannten Eigenschaften zu besitzen.

Die Regel für die Kodierung der Gruppenzugehörigkeit einer Beobachtungseinheit (Be) als Vektor der unabhängigen Variablen  $x_q$  ist dann besonders einfach. Für jede Zelle benötigen wir dann eine unabhängige Variable  $x_q$ , die folgendermaßen definiert ist:

$$x_q := \begin{cases} 1, & \text{wenn die Be zur } q\text{-ten Zelle gehört} \\ 0, & \text{andernfalls.} \end{cases}$$

Wir überzeugen uns davon, daß dann  $\beta_{qp}$  der bedingte Erwartungswert von  $y_p$  in der  $q$ -ten Zelle ist. Dazu gehen wir von der Identifikationsgleichung 2.2.15

$$B = E(x'x)^{-1} E(x'y)$$

aus. Für  $E(x'x)$  gilt in diesem Fall

$$(3.2.1) \quad E(x'x) = \begin{pmatrix} P(x_1=1) & & & & & & & O \\ & \ddots & & & & & & \\ & & \ddots & & & & & \\ & & & O & & & & P(x_Q=1) \end{pmatrix},$$

da die stochastische Variable  $x_q \cdot x_{q^*}$ ,  $q, q^* \in \{1, 2, \dots, Q\}$ , nur dann einen Wert ungleich Null annehmen kann, wenn  $q = q^*$ . In diesem Fall nimmt  $x_q \cdot x_{q^*} = x_q \cdot x_q$  den Wert Eins an, und aus der Definition des Erwartungswertes (siehe Gleichung A.1 des Anhangs A von Moosbrugger in diesem Band) folgt  $E(x_q \cdot x_q) = 1 \cdot P(x_q \cdot x_q = 1) + 0 \cdot P(x_q \cdot x_q = 0) = P(x_q = 1)$ .

Die Inverse einer solchen Diagonalmatrix ist wiederum eine Diagonalmatrix mit reziproken Werten der Ausgangsmatrix als Komponenten (vgl. z.B. Moosbrugger, 1978, S. 32).

$$(3.2.2) \quad E(x'x)^{-1} = \begin{pmatrix} \frac{1}{P(x_1=1)} & & & & & & & O \\ & \ddots & & & & & & \\ & & \ddots & & & & & \\ & & & O & & & & \frac{1}{P(x_Q=1)} \end{pmatrix}.$$

Für die Matrix  $E(x'y)$  schließlich erhalten wir

$$(3.2.3) \quad E(\mathbf{x}'\mathbf{y}) = \begin{pmatrix} E(y_1 \mid x_1=1) \cdot P(x_1=1) & \dots & E(y_p \mid x_1=1) \cdot P(x_1=1) \\ \vdots & & \vdots \\ E(y_1 \mid x_Q=1) \cdot P(x_Q=1) & \dots & E(y_p \mid x_Q=1) \cdot P(x_Q=1) \end{pmatrix},$$

wobei wir für jede Komponente dieser Matrix die Regel

$$\langle x \cdot y \rangle = \sum_{i=1}^I x_i \cdot E(y \mid x=x_i) \cdot P(x=x_i).$$

benutzen. Dabei sind  $x$  und  $y$  stochastische Variablen, wobei  $x$  jedoch diskret ist mit den 1 Realisationen  $x_i$ .

Ausmultiplizieren von  $E(\mathbf{x}'\mathbf{x})^{-1} E(\mathbf{x}'\mathbf{y})$  schließlich führt zu

$$(3.2.4) \quad \mathbf{B} = \begin{pmatrix} E(y_1 \mid x_1=1) =: \mu_{11} & \dots & E(y_p \mid x_1=1) =: \mu_{1p} \\ \vdots & & \vdots \\ E(y_1 \mid x_Q=1) =: \mu_{Q1} & \dots & E(y_p \mid x_Q=1) =: \mu_{QP} \end{pmatrix}.$$

Diese Kodierungsweise nach dem Zellenmittelwertemodell hat nicht nur den Vorteil, daß die  $\beta_{qp}$  einfach zu interpretieren sind, sondern auch die rechen-technischen Vorteile, daß nämlich ein Schätzer  $\hat{\mathbf{B}}$  von  $\mathbf{B}$  ohne faktische Kodierung der Matrix  $X$  sofort berechnet werden kann, daß die Inverse von  $X'X$  immer existiert, es sei denn, in einer Zelle liegen keine Beobachtungen vor, und daß die Inverse von  $X'X$  sehr leicht, auch ohne Rechner, zu bilden ist, indem man einfach den reziproken Wert der jeweiligen Komponente der Matrix  $X'X$  nimmt, die ja selbst eine Diagonalmatrix der Zellenfrequenzen ist.

Bei den im folgenden zu behandelnden Designs können wir also immer davon ausgehen, daß  $B$  die bedingten Erwartungswerte der  $P$  abhängigen Variablen in den  $Q$  Zellen enthält, also die Zellenmittelwerte  $\mu_{qp}$ . Die Residuen der  $P$ -dimensionalen Vektoren der abhängigen Variablen sind weder innerhalb einer solchen Zelle noch zwischen diesen Zellen voneinander abhängig (siehe Gleichung 2.3.11).

### 3.3 Die multivariate allgemeine lineare Hypothese

Bei allen in diesem Abschnitt 3 behandelten Designs lassen sich die Hypothesen über die Parameter in der Matrix B in der Form der *multivariaten allgemeinen linearen Hypothese*

$$(3.3.1) \quad H_0: C B A = A$$

formulieren, wobei C eine MxQ-Matrix mit M linear unabhängigen<sup>5)</sup> Zeilen ist, B die Q x P-Parametermatrix, A eine P x K-Matrix mit K linear unabhängigen Spalten und A eine MxK-Matrix. Die Matrizen C, A und A sind gemäß der jeweiligen Hypothese vom Anwender zu spezifizieren, wobei mit C Hypothesen über Parameter formuliert werden, die sich in jeweils einer Spalte von B befinden, und mit A Hypothesen über Parameter, die sich jeweils in einer Zeile von B befinden. Auf Einzelheiten werden wir in den nachfolgenden Abschnitten ausführlich eingehen.

Einen Zeilenvektor  $\psi_m = c_m B I$  nennen wir einen *einfachen P-varianen Vergleich der unabhängigen Variablen*. Gilt dabei für den Zeilenvektor  $c_m$  auch

$$(3.3.2) \quad c_m 1 = 0,$$

(1 ist ein passender Spalteneinheitsvektor), so sprechen wir auch von einem *Kontrast* anstatt von einem Vergleich.

Einen Spaltenvektor  $\psi_k = I B a_k$  nennen wir einen *einfachen Q-varianen Vergleich der abhängigen Variablen*, bzw. Kontrast, wenn dabei für den Spaltenvektor  $a_k$

$$(3.3.3) \quad 1' a_k = 0$$

gilt.

Mehrere solcher Zeilen- oder Spaltenvektoren heißen globaler Vergleich bzw. Kontrast.

Die Koeffizientenvektoren zweier Vergleiche bzw. Kontraste  $\psi_m$  und  $\psi_{m^*}$  oder  $\psi_k$  und  $\psi_{k^*}$  heißen *orthogonal*, wenn für ihr Skalarprodukt

$$(3.3.4) \quad c_m c'_{m^*} = 0$$

oder

$$(3.3.5) \quad a_k' a_{k^*} = 0$$

<sup>5)</sup> Endlich viele Vektoren  $v_1, \dots, v_M$  eines Vektorraums heißen linear *unabhängig*, wenn aus  $c_1 v_1 + \dots + c_M v_M = 0$  immer  $c_1 = \dots = c_M = 0$  folgt (vgl. Kowalsky, 1969, S. 33). Sind  $v_1, \dots, v_M$  linear abhängig, so lässt sich daher immer mindestens einer von ihnen als Linearkombination der anderen darstellen (vgl. z.B. auch Moosbrugger, 1978, S. 35).

gilt. Bei gleichen Zellenfrequenzen sind dann auch die Vergleiche bzw. Kontraste selbst orthogonal (d.h. unkorreliert).

Im folgenden werden wir die Spezifizierung der Gleichung 3.3.1 bei verschiedenen Designs behandeln.

### 3.4 Gekreuzte Faktoren über den unabhängigen Variablen

Wir betrachten nun ein Design, bei dem bei jeder der N Beobachtungseinheiten genau P abhängige Variablen  $y_1, \dots, y_p$  erhoben werden, und bei dem jede dieser Beachtungseinheiten in genau eine von z.B. 6 Zellen gehört, die durch das Kreuzen eines Faktors A mit zwei Stufen und eines Faktors B mit drei Stufen entstehen (siehe Tabelle 3.1). Dabei kann es sich um experimentelle Faktoren handeln, oder aber auch um solche, die man durch „reines Beobachten“<sup>6)</sup> gewinnt, wie z.B. „Geschlecht“, „Bildungsklasse“ etc.

Wir gehen zunächst von einer eher *exploratorischen* Situation aus, in der man *keine gezielten Hypothesen* hat. Bei einem solchen Design ist dann zunächst von Interesse, ob in einer Stichprobe überhaupt „signifikante“, d.h. überzufällige Mittelwerteunterschiede zwischen den 6 Zellen vorliegen. Die verbal formulierte Nullhypothese besagt dann, daß alle sechs Erwartungswertvektoren  $\beta_q = (\beta_{q1}, \dots, \beta_{qP})$ ,  $q \in \{1, 2, \dots, 6\}$ , gleich sind, oder mit anderen Worten, daß weder ein Haupteffekt von Faktor A, noch ein Haupteffekt von Faktor B noch ein Wechselwirkungseffekt der Faktorstufenkombinationen AB besteht. Die Bedeutung dieser Begriffe erläutert z.B. Moosbrugger (1982) in diesem Band, S. 18-19.

Prinzipiell lassen sich nun viele verschiedene Möglichkeiten denken, die *globale Hypothese*, daß keine Unterschiede zwischen den sechs Zeilenvektoren  $\beta_q$  bestehen, in der Form der multivariaten allgemeinen linearen Hypothese zu formulieren. Diese unterscheiden sich jedoch nicht in den resultierenden Quadratsummenmatrizen, mit denen man die Hypothesenbewertung durchführt (siehe Abschnitt 5).

Eine der einfachsten Möglichkeiten besteht darin, die Hypothese so zu formulieren, daß alle benachbarten Zeilenvektoren  $\beta_q$  bzw.  $\mu_q$  gleich sind, d.h. daß deren Differenz jeweils gleich dem Nullvektor ist:

$$(3.4.1) \quad H_{0_{A,B,AB}}: \quad \begin{aligned} \mu_1 - \mu_2 &= 0 \\ \mu_2 - \mu_3 &= 0 \\ \mu_3 - \mu_4 &= 0 \\ \mu_4 - \mu_5 &= 0 \\ \mu_5 - \mu_6 &= 0. \end{aligned}$$

<sup>6)</sup> Edwards (1971, S. 8f.) führt in dieser Kategorie z.B. „organismische“ Variablen an.

Tabelle 3.1: Gekreuzte Faktoren A und B als Gliederung der Q unabhängigen Variablen  $x_q$ , welche hier an insgesamt  $N = N_1 + \dots + N_6$  Beobachtungseinheiten erhoben wurden. Beispiel für Faktor A: Geschlecht ( $A_1, A_2$ ), für Faktor B: Altersstufen ( $B_1, B_2, B_3$ ). Das Datenschema zeigt die  $P = 4$  abhängigen Variablen  $y_p$  pro Beobachtungseinheit und deren bedingte Erwartungswerte  $\hat{\mu}_{qp}$ .

Faktor A	Faktor B	Zellen- nummer q	Variablennummer p			
			1	2	3	4
A <sub>1</sub>	B <sub>1</sub>	1	$y_{111}$ ⋮ $y_{N_111}$	$y_{112}$ ⋮ $y_{N_112}$	$y_{113}$ ⋮ $y_{N_113}$	$y_{114}$ ⋮ $y_{N_114}$
			$\hat{\mu}_{11}$	$\hat{\mu}_{12}$	$\hat{\mu}_{13}$	$\hat{\mu}_{14}$
			$y_{121}$ ⋮ $y_{N_21}$	$y_{122}$ ⋮ $y_{N_22}$	$y_{123}$ ⋮ $y_{N_223}$	$y_{124}$ ⋮ $y_{N_224}$
			$\hat{\mu}_{21}$	$\hat{\mu}_{22}$	$\hat{\mu}_{23}$	$\hat{\mu}_{24}$
	B <sub>3</sub>	3	$y_{131}$ ⋮ $y_{N_31}$	$y_{132}$ ⋮ $y_{N_32}$	$y_{133}$ ⋮ $y_{N_33}$	$y_{134}$ ⋮ $y_{N_34}$
			$\hat{\mu}_{31}$	$\hat{\mu}_{32}$	$\hat{\mu}_{33}$	$\hat{\mu}_{34}$
			$y_{141}$ ⋮ $y_{N_41}$	$y_{142}$ ⋮ $y_{N_42}$	$y_{143}$ ⋮ $y_{N_43}$	$y_{144}$ ⋮ $y_{N_44}$
			$\hat{\mu}_{41}$	$\hat{\mu}_{42}$	$\hat{\mu}_{43}$	$\hat{\mu}_{44}$
	B <sub>2</sub>	5	$y_{151}$ ⋮ $y_{N_51}$	$y_{152}$ ⋮ $y_{N_52}$	$y_{153}$ ⋮ $y_{N_53}$	$y_{154}$ ⋮ $y_{N_54}$
			$\hat{\mu}_{51}$	$\hat{\mu}_{52}$	$\hat{\mu}_{53}$	$\hat{\mu}_{54}$
			$y_{161}$ ⋮ $y_{N_61}$	$y_{162}$ ⋮ $y_{N_62}$	$y_{163}$ ⋮ $y_{N_63}$	$y_{164}$ ⋮ $y_{N_64}$
			$\hat{\mu}_{61}$	$\hat{\mu}_{62}$	$\hat{\mu}_{63}$	$\hat{\mu}_{64}$

Gliederung der unabhängigen Variablen

Da gemäß Gleichung 3.2.4

$$(3.4.2) \quad \mathbf{B} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \\ \vdots \\ \boldsymbol{\mu}_6 \end{pmatrix}$$

gilt, wählen wir die folgende Hypothesenmatrix

$$(3.4.3) \quad \mathbf{C}_{A,B,AB} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix},$$

wobei wir, falls nicht explizit anders angegeben, gleiche Zellenfrequenzen voraussetzen.

Die Matrix A in der allgemeinen linearen Hypothese  $C B A = A$  erlaubt, die P abhängigen Variablen beliebig linear zusammenzufassen. Man kann damit praktisch neue abhängige Variablen bilden. In anderen Designs kann A jedoch auch die „Hauptrolle“ beim Formulieren von Hypothesen spielen, wie wir weiter unten zeigen werden.

$A_{A,B,AB}$  sei hier die PxP-Einheitsmatrix I, und  $\Delta_{A,B,AB}$  eine 5xP-Nullmatrix. Ausmultiplizieren von  $C_{A,B,AB} B A_{A,B,AB} = \Delta_{A,B,AB}$  führt dann zu den Gleichungen 3.4.1. Die so formulierte Hypothese lässt sich mit den in Abschnitt 5 angegebenen Verfahren bewerten. Falls sich dabei ergibt, daß man die Nullhypothese beibehalten sollte, kann man das Verfahren hier abbrechen. Andernfalls jedoch kann man mit der nächsten Hypothese fortfahren.

Es ist dabei u.E. zu empfehlen, als nächstes die Nullhypothese zu untersuchen, daß keine *Interaktion (Wechselwirkung)* zwischen den Faktoren A und B bestehen, nämlich, daß die Unterschiede zwischen den Erwartungswertvektoren der Stufen  $A_1$  und  $A_2$  auf allen Stufen von Faktor B gleich sind:

$$(3.4.4) \quad H_{0_{AB}}: \begin{aligned} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_4) - (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_5) &= 0 \\ (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_5) - (\boldsymbol{\mu}_3 - \boldsymbol{\mu}_6) &= 0. \end{aligned}$$

In der Form  $C B A = A$  wählen wir

$$(3.4.5) \quad \mathbf{C}_{AB} = \begin{pmatrix} 1 & -1 & 0 & -1 & 1 & 0 \\ 0 & 1 & -1 & 0 & -1 & 1 \end{pmatrix},$$

$A_{AB} = 1$  und  $\Delta_{AB} = 0$ . Multipliziert man  $C_{AB} B A_{AB} = \Delta_{AB}$  aus, so erhält man die Gleichungen 3.4.4. Die Bewertung führt man wiederum mit den in Abschnitt 5 angegebenen Verfahren durch.

Sollte man zu dem Ergebnis kommen, daß wohl keine Interaktionen vorliegen, so kann man die *Haupteffekte* der Faktoren A bzw. B mit den Nullhypotesen untersuchen, daß erstens die Summen der Erwartungswertevektoren auf den Stufen von Faktor A bzw. zweitens auf den Stufen von Faktor B gleich sind. Es ist u.E. auch dann sinnvoll, diese Hypothesen zu untersuchen, wenn eine Interaktion AxB besteht, falls man am durchschnittlichen Effekt z.B. des Faktors A über die Stufen von B hinweg (bzw. umgekehrt) interessiert ist.

Die Nullhypothese für Faktor A kann man so formulieren:

$$(3.4.6) \quad H_{0_A}: (\mu_1 + \mu_2 + \mu_3) - (\mu_4 + \mu_5 + \mu_6) = 0.$$

in der Form  $C B A = A$  wählen wir

$$(3.4.7) \quad C_A = [1 \quad 1 \quad 1 \quad -1 \quad -1 \quad -1],$$

sowie  $A_A = I$  und  $\Delta_A = 0$ .

Die Nullhypothese für Faktor B ist

$$(3.4.8) \quad H_{0_B}: \begin{aligned} &(\mu_1 + \mu_4) - (\mu_2 + \mu_5) = 0 \\ &(\mu_2 + \mu_5) - (\mu_3 + \mu_6) = 0 \end{aligned},$$

und in der Form  $C B A = A$  ist dann

$$(3.4.9) \quad C_B = \begin{pmatrix} 1 & -1 & 0 & 1 & -1 & 0 \\ 0 & 1 & -1 & 0 & 1 & -1 \end{pmatrix},$$

$A_B = I$  und  $\Delta_B = 0$  zu wählen. Die Hypothesen für die Haupteffekte (vgl. auch Moosbrugger, 1978, S. 139-143) lassen sich mit den Verfahren bewerten, die in Abschnitt 5 angegeben sind.

Wie ist nun die Hypothesenmatrix C für den Fall zu wählen, daß man z.B. den „Effekt“ von Faktor A auf der Stufe 1 des Faktors B untersuchen will? Dazu wäre einfach

$$(3.4.10) \quad C_{A|B_1} = [1 \quad 0 \quad 0 \quad -1 \quad 0 \quad 0]$$

zu setzen, was zusammen mit  $A_{A|B_1} = I$  und  $\Delta_{A|B_1} = 0$  der Nullhypothese

$$H_{0_{A|B_1}}: \mu_1 - \mu_4 = 0$$

entspricht.

Man beachte, daß jede Zeile der Matrix C einer eigenen (Sub-) Hypothese entspricht, die man nach den gleichen allgemeinen Methoden des Abschnitts 5 untersuchen kann. Solche gezielten Einzelvergleiche sind legitim, falls gezielte Hypothesen vor der Datenerhebung bestehen<sup>7)</sup>.

Die Verallgemeinerung des hier behandelten Designs auf beliebig viele Faktoren mit beliebig vielen Stufen sollte keine Schwierigkeiten bereiten. Bei der Anordnung der Daten ist lediglich darauf zu achten, daß die Beobachtungseinheiten n untereinander angeordnet werden, sofern man bereit ist, die Unabhängigkeitssannahme (2.3.11) zu machen.

Es wurde bereits weiter oben darauf hingewiesen, daß die hier gewählten Hypothesen nicht die einzigen möglichen und richtigen sind. Man könnte z.B. an der Frage interessiert sein, ob der Durchschnitt der Erwartungswerte in den Zellen 1, 2, 4 und 5 gleich dem Durchschnitt der Erwartungswerte in den Zellen 3 und 6 ist. Allgemein sind auch Hypothesen der Form  $C B A = A \neq 0$  möglich.

Zum Abschluß dieses Abschnitts sei davor gewarnt, zu viele Einzelvergleiche durchzuführen. Dies führt zu einer Akkumulation des a-Fehlers. Wenn bei einem Signifikanztest die Irrtumswahrscheinlichkeit für die fälschliche Ablehnung der Nullhypothese  $\alpha = 0.05$  beträgt, so liegt sie bei fünf unabhängigen Signifikanztests auf diesem Niveau schon bei  $\alpha_5 = 1 - (1 - \alpha)^5 = 1 - 0.77$ , d.h. die Wahrscheinlichkeit bei fünf unabhängigen Signifikanztests fälschlicherweise einen „signifikanten“ zu finden, beträgt  $\alpha_5 = 0.23$ . Sind viele Signifikanztests durchzuführen, sollte man daher auf die Verfahren „simultane Mittelwertvergleiche“ zurückgreifen (vgl. Aitkin, 1969, z.B. Miller, 1966, oder Gabriel, 1968, 1969a, b).

### 3.5 Gekreuzte Faktoren über den abhängigen Variablen

Wir betrachten nun die P abhängigen Variablen genauer und nehmen an, daß sie sich ebenfalls faktoriell gliedern lassen (siehe Tabelle 3.2), so, wie sich die Q unabhängigen Gruppen im vorangegangenen Abschnitt strukturieren ließen. Die Aufteilung der Gruppen wollen wir hier zunächst nicht weiter beachten, da wir sie ja bereits im vorangegangenen Abschnitt behandelt haben.

<sup>7)</sup> Andernfalls handelt es sich nicht um Hypothesentesten, sondern um die Exploration von Daten, und die dabei gefundenen „signifikanten“ Tests sollte man nicht als signifikant, sondern als statistisch auffällig bezeichnen, da die Irrtumswahrscheinlichkeitsangaben (a-Niveau) dann sehr irreführend sein können.

Tabelle 3.2: *Gekreuzte Faktoren U und V als Gliederung der P abhängigen Variablen  $y_p$ , welche an insgesamt  $N = N_1 + \dots + N_6$  Beobachtungseinheiten (Ben) erhoben wurden. Dabei wird die Unabhängigkeitssannahme, die in Gleichung 2.3.11 enthalten ist, vorausgesetzt. Beispiel für Faktor U: Schulleistungstests ( $U_1, U_2$ ), für Faktor V: Vortest ( $V_1$ ), Nachtest ( $V_2$ ). Das Datenschema zeigt die  $P = 4$  abhängigen Variablen  $y_p$  pro Be und deren bedingte Erwartungswerte  $\hat{\mu}_{qp}$ .*

Gliederung der abhängigen Variablen

Zellennummer q	$U_1$		$U_2$		Faktor U Faktor V Variablennummer p
	$V_1$	$V_2$	$V_1$	$V_2$	
	1	2	3	4	
1	$y_{111}$ ⋮ $y_{N_111}$	$y_{112}$ ⋮ $y_{N_112}$	$y_{113}$ ⋮ $y_{N_113}$	$y_{114}$ ⋮ $y_{N_114}$	
	$\hat{\mu}_{11}$	$\hat{\mu}_{12}$	$\hat{\mu}_{13}$	$\hat{\mu}_{14}$	
	$y_{121}$ ⋮ $y_{N_211}$	$y_{122}$ ⋮ $y_{N_212}$	$y_{123}$ ⋮ $y_{N_213}$	$y_{124}$ ⋮ $y_{N_214}$	
	$\hat{\mu}_{21}$	$\hat{\mu}_{22}$	$\hat{\mu}_{23}$	$\hat{\mu}_{24}$	
3	$y_{131}$ ⋮ $y_{N_311}$	$y_{132}$ ⋮ $y_{N_312}$	$y_{133}$ ⋮ $y_{N_313}$	$y_{134}$ ⋮ $y_{N_314}$	
	$\hat{\mu}_{31}$	$\hat{\mu}_{32}$	$\hat{\mu}_{33}$	$\hat{\mu}_{34}$	
	$y_{141}$ ⋮ $y_{N_411}$	$y_{142}$ ⋮ $y_{N_412}$	$y_{143}$ ⋮ $y_{N_413}$	$y_{144}$ ⋮ $y_{N_414}$	
	$\hat{\mu}_{41}$	$\hat{\mu}_{42}$	$\hat{\mu}_{43}$	$\hat{\mu}_{44}$	
5	$y_{151}$ ⋮ $y_{N_511}$	$y_{152}$ ⋮ $y_{N_512}$	$y_{153}$ ⋮ $y_{N_513}$	$y_{154}$ ⋮ $y_{N_514}$	
	$\hat{\mu}_{51}$	$\hat{\mu}_{52}$	$\hat{\mu}_{53}$	$\hat{\mu}_{54}$	
	$y_{161}$ ⋮ $y_{N_611}$	$y_{162}$ ⋮ $y_{N_612}$	$y_{163}$ ⋮ $y_{N_613}$	$y_{164}$ ⋮ $y_{N_614}$	
	$\hat{\mu}_{61}$	$\hat{\mu}_{62}$	$\hat{\mu}_{63}$	$\hat{\mu}_{64}$	

Die Faktoren U und V können in verschiedenen Forschungssituationen entstehen: Falls man nicht bereit ist, die in Gleichung 2.3.11 enthaltene Unabhängigkeitsannahme zu machen, kann z.B.  $V_1$  Vor- und  $V_2$  Nachtest von 2 inhaltlich verschiedenen abhängigen Variablen ( $U_1, U_2$ ) sein, oder umgekehrt können  $U_1$  und  $U_2$  wiederholte Messungen von 2 inhaltlich verschiedenen abhängigen Variablen sein ( $V_1$  und  $V_2$ ). Genausogut kann es sich jedoch auch bei  $U_1$  und  $U_2$  und bei  $V_1$  und  $V_2$  um wiederholte Messungen handeln, wobei  $U_1, U_2$  aufeinanderfolgende experimentelle Versuchsbedingungen sind. Kann man hingegen die Annahme  $C(\boldsymbol{\varepsilon}_n, \boldsymbol{\varepsilon}_{n^*}) = 0$  bei  $n \neq n^*$ , für die Residuenvektoren  $\boldsymbol{\varepsilon}_n = y_n - E(y_n | x_n)$  machen, wobei  $y_n$  und  $y_{n^*}$  die abhängigen Variablen derselben Person zu verschiedenen Zeitpunkten enthält, so würde man besser die Meßwiederholungen „untereinander“ anordnen, wodurch der Index n dann nicht mehr die Person, sondern eine Person-Zeit-Kombination bezeichnet. Wottawa (1981) hat ein Verfahren angegeben, wie die Annahme der Unabhängigkeit der Residuen überprüft werden kann.

Formal betrachtet haben wir von jeder Beobachtungseinheit hier insgesamt  $P = 2 \cdot 2$  abhängige Variablen vorliegen. Die Gleichungen 2.3.8 bis 2.3.11 werden also weiterhin vorausgesetzt, wobei hier  $N = N_1 + \dots + N_4$  gilt, und  $\Sigma$  die Dimension  $P \times P = 4 \times 4$  hat.

Da wir auch hier ein Zellenmittelwertmodell zugrundelegen, sind die Parameter in der Matrix B wiederum die Erwartungswerte der  $P=4$  abhängigen Variablen in den  $Q=6$  Zellen, deren Anordnung in Tabelle 3.2 wiedergegeben ist.

In formaler Hinsicht unterscheidet sich das Design der Tabelle 3.2 nicht von dem der Tabelle 3.1. Der einzige Unterschied besteht darin, daß hier die  $P = 2 \cdot 2$  **abhängigen** Variablen untergliedert wurden, wobei die Untergliederungen der Q Gruppen nach den Faktoren A und B, die wir in Tabelle 3.1 vorgenommen hatten, außer acht gelassen wird.

Die Hypothesen über die Faktoren U, V und deren Interaktion  $U \times V$  können nach der gleichen Methode wie im vorangegangenen Abschnitt formuliert werden. Der einzige Unterschied besteht darin, daß hier die Matrix A der multivariaten allgemeinen linearen Hypothese  $C B A = A$  Verwendung findet und C jeweils die  $Q \times Q$ -Einheitsmatrix ist. Die einzelnen Hypothesenmatrizen A können dabei so gewählt werden, wie die Hypothesenmatrizen C der entsprechenden Hypothesen in Abschnitt 3.4, aber in transponierter Form.

Bei einem Design wie dem vorliegenden können zusätzlich noch Interaktionen zwischen den „unabhängigen Faktoren“ A und B (entsprechend Tabelle 3.1) und den „abhängigen Faktoren“ U und V (entsprechend Tabelle 3.2) geprüft werden. So kann man z.B. die Nullhypothese formulieren, daß keine Interaktion zwischen den Q Gruppen und den P abhängigen Variablen bestehen:

(3.5.1)

 $H_{0_{(A,B,AB)} \times (U,V,UV) } :$ 

$$\begin{aligned} (\mu_{11} - \mu_{12}) &= (\mu_{21} - \mu_{22}) = \dots = (\mu_{61} - \mu_{62}) \\ (\mu_{12} - \mu_{13}) &= (\mu_{22} - \mu_{23}) = \dots = (\mu_{62} - \mu_{63}) \\ (\mu_{13} - \mu_{14}) &= (\mu_{23} - \mu_{24}) = \dots = (\mu_{63} - \mu_{64}). \end{aligned}$$

Anschaulich bedeutet das, daß die Differenzen zwischen den Erwartungswerten der P=4 abhängigen Variablen von den Q=6 Zellen nicht moderiert werden (vgl. Moosbrugger, 1982, in diesem Band, S. 19). Dabei können diese Differenzen zwischen den P=4 Variablen wohl verschieden sein. Diese Nullhypothese können wir auch in Form von 15 Differenzen anschreiben, welche Null sein müssen:

(3.5.2)

 $H_{0_{(A,B,AB)} \times (U,V,UV) } :$ 

$$\begin{aligned} (\mu_{11} - \mu_{12}) - (\mu_{21} - \mu_{22}) &= (\mu_{12} - \mu_{13}) - (\mu_{22} - \mu_{23}) = (\mu_{13} - \mu_{14}) - (\mu_{23} - \mu_{24}) = 0 \\ (\mu_{21} - \mu_{22}) - (\mu_{31} - \mu_{32}) &= (\mu_{22} - \mu_{23}) - (\mu_{32} - \mu_{33}) = (\mu_{23} - \mu_{24}) - (\mu_{33} - \mu_{34}) = 0 \\ &\vdots && \vdots && \vdots \\ (\mu_{51} - \mu_{52}) - (\mu_{61} - \mu_{62}) &= (\mu_{52} - \mu_{53}) - (\mu_{62} - \mu_{63}) = (\mu_{53} - \mu_{54}) - (\mu_{63} - \mu_{64}) = 0 \end{aligned}$$

Zur Formulierung dieser Wechselwirkungshypothese zwischen den unabhängigen und abhängigen Faktoren in der Form C B A = A wählen wir

(3.5.3)

$$C_{A,B,AB} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

und zugleich

(3.5.4)

$$A_{U,V,UV} = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{pmatrix}$$

sowie  $\Delta_{(A,B,AB) \times (U,V,UV)} = 0$ . Ebenso gut ließen sich aber auch Hypothesen formulieren, daß keine Interaktionen A x U, A x V etc. bestehen. Nachdem einmal die Arbeit der Formulierung der Hypothesen in der Form C B A = A geleistet ist, können diese nach den Verfahren untersucht werden, die wir in Abschnitt 5 behandeln.

Die Verallgemeinerung eines solchen Designs auch für beliebig viele Faktoren über den abhängigen Variablen ist offensichtlich. Bei der Festlegung des Designs ist lediglich darauf zu achten, daß für jede Be (z.B. Person oder Person-Zeit-Kombination) die P abhängigen Variablen jeweils nebeneinander und die Beobachtungseinheiten untereinander angeordnet werden, so daß die Gleichungen 2.3.6 bis 2.3.11 erfüllt sind.

### 3.6 Hierarchische Faktoren über den unabhängigen Variablen

Wir betrachten nun ein Design, bei dem ein fünfstufiger Faktor B einem zweistufigen Faktor A hierarchisch untergeordnet ist, bei dem also jede Stufe von B nur mit einer Stufe des Faktors A zusammen vorkommt. Beide Faktoren strukturieren dabei die  $Q=5$  Zellen, innerhalb derer sich dann insgesamt  $N = N_1 + \dots + N_5$  P-dimensionale Vektoren abhängiger Variablen befinden. Ein typischer Fall wäre, daß man P abhängige Variablen bei zwei verschiedenen Lehrmethoden (Faktor A) untersuchen will, wobei man die Schulklassen B<sub>1</sub> bis B<sub>5</sub> einbezieht.

Wir setzen wieder voraus, daß die Zellenmittelwertekodierung für X angewandt wurde, so daß B die  $5 \times P$ -Matrix der bedingten Erwartungswerte der abhängigen Variablen in den fünf Zellen ist, und zwar in der gleichen Anordnung wie diese in Tabelle 3.3 dargestellt sind.

Bei einem solchen Design sind mehrere Hypothesen von Interesse. Eine erste globale Hypothese besagt, daß keine Unterschiede zwischen den Gruppen B<sub>1</sub> bis B<sub>5</sub> bezüglich der Erwartungswerte der P abhängigen Variablen bestehen.

Diese Hypothese läßt sich in der Form der allgemeinen multivariaten linearen Hypothese C B A = A darstellen, indem wir z.B.

$$(3.6.1) \quad \mathbf{C}_{A,B} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

$A_{A,B} = I$  und  $\Delta_{A,B} = 0$  wählen. Falls die Bewertung nach den in Abschnitt 5 angegebenen Verfahren ergibt, daß diese Hypothese beizubehalten sei, kann das Verfahren abgebrochen werden. Andernfalls überprüfen wir als zweite Hypothese, daß sich die Erwartungswerte der P abhängigen Variablen zwischen den Gruppen A<sub>1</sub> und A<sub>2</sub> nicht unterscheiden. Diese läßt sich in der Form C B A = A formulieren, indem wir z.B.

$$(3.6.2) \quad \mathbf{C}_A = [1/2 \quad 1/2 \quad -1/3 \quad -1/3 \quad -1/3],$$

$A_A = I$  und  $\Delta_A = 0$  wählen.

Der Faktor B ist unter Faktor A „genestet“. Folglich können Hypothesen über Faktor B nur auf den Stufen von Faktor A geprüft werden. Wir formulieren die dritte bzw. vierte Hypothese, daß sich die Erwartungswerte der P abhängigen Variablen zwischen den Gruppen B<sub>1</sub> und B<sub>2</sub> bzw. B<sub>3</sub> bis B<sub>5</sub> auf der Stufe A<sub>1</sub> bzw. A<sub>2</sub> nicht unterscheiden, in der Form C B A = A, indem wir z.B.

Tabelle 3.3 Hierarchische Faktoren A und B als Gliederung der Q unabhängigen Variablen  $x_q$ , welche an insgesamt  $N = N_1 + \dots + N_5$  Beobachtungseinheiten erhoben wurden. Beispiel für Faktor A: Lehrmethoden ( $A_1, A_2$ ), für Faktor B: Schulklassen ( $B_1$  bis  $B_5$ ). Faktor B ist dem Faktor A untergeordnet („genestet“), weil jede Stufe von B nur unter einer Stufe von A vorkommt. Das Datenschema zeigt die  $P=5$  abhängigen Variablen  $y_p$  pro Beobachtungseinheit und deren bedingte Erwartungswerte  $\hat{y}_{qp}$ .

			Variablennummer p				
		Zellennummer q	1	2	3	4	5
Faktor A	Faktor B		$y_{111}$ ⋮ $y_{N_111}$	$y_{112}$ ⋮ $y_{N_112}$	$y_{113}$ ⋮ $y_{N_113}$	$y_{114}$ ⋮ $y_{N_114}$	$y_{115}$ ⋮ $y_{N_115}$
A <sub>1</sub>	B <sub>1</sub>	1	$\hat{y}_{11}$	$\hat{y}_{12}$	$\hat{y}_{13}$	$\hat{y}_{14}$	$\hat{y}_{15}$
			$y_{121}$ ⋮ $y_{N_211}$	$y_{122}$ ⋮ $y_{N_212}$	$y_{123}$ ⋮ $y_{N_213}$	$y_{124}$ ⋮ $y_{N_214}$	$y_{125}$ ⋮ $y_{N_215}$
		2	$\hat{y}_{21}$	$\hat{y}_{22}$	$\hat{y}_{23}$	$\hat{y}_{24}$	$\hat{y}_{25}$
			$y_{131}$ ⋮ $y_{N_311}$	$y_{132}$ ⋮ $y_{N_312}$	$y_{133}$ ⋮ $y_{N_313}$	$y_{134}$ ⋮ $y_{N_314}$	$y_{135}$ ⋮ $y_{N_315}$
			$\hat{y}_{31}$	$\hat{y}_{32}$	$\hat{y}_{33}$	$\hat{y}_{34}$	$\hat{y}_{35}$
	B <sub>2</sub>	3	$y_{141}$ ⋮ $y_{N_411}$	$y_{142}$ ⋮ $y_{N_412}$	$y_{143}$ ⋮ $y_{N_413}$	$y_{144}$ ⋮ $y_{N_414}$	$y_{145}$ ⋮ $y_{N_415}$
			$\hat{y}_{41}$	$\hat{y}_{42}$	$\hat{y}_{43}$	$\hat{y}_{44}$	$\hat{y}_{45}$
		4	$y_{151}$ ⋮ $y_{N_511}$	$y_{152}$ ⋮ $y_{N_512}$	$y_{153}$ ⋮ $y_{N_513}$	$y_{154}$ ⋮ $y_{N_514}$	$y_{155}$ ⋮ $y_{N_515}$
			$\hat{y}_{51}$	$\hat{y}_{52}$	$\hat{y}_{53}$	$\hat{y}_{54}$	$\hat{y}_{55}$

Gliederung der unabhängigen Variablen

$$(3.6.3) \quad \mathbf{C}_{B|A_1} = [ \begin{array}{cccccc} 1 & -1 & 0 & 0 & 0 \end{array} ]$$

bei  $\mathbf{A}_{B|A} = I$  und  $\Delta_{B|A} = 0$  bzw.

$$(3.6.4) \quad \mathbf{C}_{B|A_2} = \left[ \begin{array}{ccccc} 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{array} \right]$$

bei  $\mathbf{A}_{B|A} = I$  und  $\Delta_{B|} = 0$  wählen.

Die beiden Hypothesen über Faktor B können auch gemeinsam formuliert und überprüft werden.

Interaktionen zwischen den Faktoren A und B lassen sich bei einem hierarchischen Design nicht überprüfen.

### 3.7 Hierarchische Faktoren über den abhängigen Variablen

Das im vorangegangenen Abschnitt dargestellte Prinzip der Strukturierung der Beobachtungseinheiten durch hierarchische Faktoren lässt sich auch auf die abhängigen Variablen übertragen. So ist z.B. denkbar, daß mit jeder Person fünf Versuchsphasen ( $V_1$  bis  $V_5$ ) durchgeführt werden, die sich aber in zwei Hauptphasen ( $U_1$  und  $U_2$ ) aufgliedern lassen, wobei  $V_1$  bis  $V_3$  der Stufe  $U_1$  untergeordnet sind und  $V_4$  bis  $V_5$  der Stufe  $U_2$ . Sind wir aus theoretischen Gründen (z.B. wegen Existenz von „Lerneffekten“) nicht in der Lage, die Gleichung 2.3.11 vorauszusetzen, wenn  $n$  eine Person-Zeit-Kombination indiziert, so können wir die fünf Meßwiederholungen als Realisationen von fünf (formal verschiedenen) abhängigen Variablen auffassen, so daß  $n$  nun die Person angibt, von der jeweils 5 Meßwerte vorliegen. So umstrukturiert wären die Gleichungen 2.3.8 bis 2.3.11 wieder gültig, wenn Abhängigkeiten zwischen den Personen ausgeschlossen werden können.

In einer ersten globalen Nullhypothese behaupten wir, daß sich die Erwartungswerte der abhängigen Variablen zwischen den fünf Versuchsphasen nicht unterscheiden. Dazu würde man  $C_{U,V} = I$ ,  $\Delta_{U,V} = 0$  und

$$(3.7.1) \quad \mathbf{A}_{U,V} = \left[ \begin{array}{cccc} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 \end{array} \right]$$

wählen. Dabei können sehr wohl Unterschiede zwischen den Q Zellen bestehen.

Tabelle 3.4: Hierarchische Faktoren U und V als Gliederung der P abhängigen Variablen  $y_p$ , welche hier an insgesamt  $N = N_1 + \dots + N_5$  Beobachtungseinheiten erhoben wurden. Beispiel für Faktor U: vormittags ( $U_1$ ), nachmittags ( $U_2$ ), für Faktor V: 5 Versuchsphasen ( $V_1$  bis  $V_5$ ). Faktor V ist Faktor U hierarchisch untergeordnet („genestet“), weil jede Stufe von V nur unter einer Stufe von U vorkommt. Das Datenschema zeigt die  $P = 5$  abhängigen Variablen  $y_p$  pro Beobachtungseinheit und deren bedingte Erwartungswerte  $\hat{\mu}_{qp}$ .

Gliederung der abhängigen Variablen

Zellennummer q	$U_1$			$U_2$		Faktor U Faktor V Variablennummer p
	$V_1$	$V_2$	$V_3$	$V_4$	$V_5$	
	1	2	3	4	5	
1	$y_{111}$ ⋮ $y_{N_111}$	$y_{112}$ ⋮ $y_{N_112}$	$y_{113}$ ⋮ $y_{N_113}$	$y_{114}$ ⋮ $y_{N_114}$	$y_{115}$ ⋮ $y_{N_115}$	
	$\hat{\mu}_{11}$	$\hat{\mu}_{12}$	$\hat{\mu}_{13}$	$\hat{\mu}_{14}$	$\hat{\mu}_{15}$	
	$y_{121}$ ⋮ $y_{N_211}$	$y_{122}$ ⋮ $y_{N_212}$	$y_{123}$ ⋮ $y_{N_213}$	$y_{124}$ ⋮ $y_{N_214}$	$y_{125}$ ⋮ $y_{N_215}$	
	$\hat{\mu}_{21}$	$\hat{\mu}_{22}$	$\hat{\mu}_{23}$	$\hat{\mu}_{24}$	$\hat{\mu}_{25}$	
	$y_{131}$ ⋮ $y_{N_311}$	$y_{132}$ ⋮ $y_{N_312}$	$y_{133}$ ⋮ $y_{N_313}$	$y_{134}$ ⋮ $y_{N_314}$	$y_{135}$ ⋮ $y_{N_315}$	
3	$\hat{\mu}_{31}$	$\hat{\mu}_{32}$	$\hat{\mu}_{33}$	$\hat{\mu}_{34}$	$\hat{\mu}_{35}$	
	$y_{141}$ ⋮ $y_{N_411}$	$y_{142}$ ⋮ $y_{N_412}$	$y_{143}$ ⋮ $y_{N_413}$	$y_{144}$ ⋮ $y_{N_414}$	$y_{145}$ ⋮ $y_{N_415}$	
	$\hat{\mu}_{41}$	$\hat{\mu}_{42}$	$\hat{\mu}_{43}$	$\hat{\mu}_{44}$	$\hat{\mu}_{45}$	
	$y_{151}$ ⋮ $y_{N_511}$	$y_{152}$ ⋮ $y_{N_512}$	$y_{153}$ ⋮ $y_{N_513}$	$y_{154}$ ⋮ $y_{N_514}$	$y_{155}$ ⋮ $y_{N_515}$	
	$\hat{\mu}_{51}$	$\hat{\mu}_{52}$	$\hat{\mu}_{53}$	$\hat{\mu}_{54}$	$\hat{\mu}_{55}$	

Als zweite Hypothese prüfen wir, ob sich die Erwartungswerte der abhängigen Variablen zwischen den Faktorstufen  $U_1$  und  $U_2$  nicht unterscheiden. Dazu wählen wir  $C_U = I$ ,  $\Delta_U = 0$  und

$$(3.7.2) \quad A_U = \begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \\ -1/2 \\ -1/2 \end{pmatrix}.$$

Der Faktor V ist unter den Faktor U „genestet“. Hypothesen über Faktor V sind folglich nur auf den Stufen von Faktor U überprüfbar:

Wir formulieren die dritte bzw. vierte Hypothese, daß sich die Erwartungswerte der P abhängigen Variablen zwischen den Stufen  $V_1$  bis  $V_3$  auf Stufe  $U_1$  bzw.  $V_4$  und  $V_5$  auf Stufe  $U_2$  nicht unterscheiden, in der Form  $C_B A = \Delta$ , indem wir z.B.

$$(3.7.3) \quad A_{V|U_1} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

bei  $C_{V|U} = I$  und  $\Delta_{V|U_1} = 0$  bzw.

$$(3.7.4) \quad A_{V|U_2} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}$$

bei  $C_{V|U} = I$  und  $\Delta_{V|U_2} = 0$  wählen.

Die beiden Hypothesen über Faktor V können auch gemeinsam formuliert und überprüft werden. Hypothesen über eine Interaktion  $U \times V$  lassen sich bei einem solchen Design nicht prüfen. Jedoch können durchaus Hypothesen über die Interaktion von „abhängigen“ und „unabhängigen“ Faktoren geprüft werden (vgl. Abschnitt 3.5).

Es sei noch bemerkt, daß die Q Zellen beliebig, d.h. kreuzfaktoriell, hierarchisch oder sonstwie, strukturiert sein können. Eine weitere Verallgemeinerung dieses Designs liegt in der Möglichkeit, zu jeder Versuchsphase mehrere abhängige Variablen zugleich zu erheben, so daß dieser neue Faktor mit U und mit V gekreuzt wäre. Prinzipiell läßt sich jede Strukturierung der abhängigen Variablen mit jeder Strukturierung der unabhängigen Beobachtungseinheiten kombinieren, womit jeweils ein neues Design konstruiert wäre.

### 3.8 Lateinisches Quadrat über den unabhängigen Variablen

Eine weitere Möglichkeit, die unabhängigen Beobachtungseinheiten zu strukturieren, ist das Prinzip des „lateinischen Quadrats“. Bei diesem Designtyp hat man mehrere Faktoren mit gleicher Stufenzahl und P abhängige Variablen vorliegen, die ihrerseits durch ein beliebiges Design strukturiert sein können.

Tabelle 3.5: Faktoren A, B und C im „lateinischen Quadrat“ als Gliederung der Q unabhängigen Variablen  $x_q$ , welche insgesamt an  $N = N_1 + \dots + N_4$  Beobachtungseinheiten erhoben wurden. Beispiel für Faktor A: Geschlecht ( $A_1, A_2$ ), für Faktor B: Sozialstatus ( $B_1, B_2$ ), Faktor C: Unterrichtsmethoden ( $C_1, C_2$ ). Das Datenschema zeigt die  $P = 4$  abhängigen Variablen  $y_p$  pro Beobachtungseinheit und deren bedingte Erwartungswerte  $\hat{\mu}_{qp}$ .

	Faktor A	Faktor B	Faktor C	Zellen- nummer q	Variablennummer p			
					1	2	3	4
Gliederung der unabhängigen Variablen	A <sub>1</sub>	B <sub>1</sub>	C <sub>1</sub>	1	$y_{111}$ ⋮ $y_{N_111}$	$y_{112}$ ⋮ $y_{N_112}$	$y_{113}$ ⋮ $y_{N_113}$	$y_{114}$ ⋮ $y_{N_114}$
					$\hat{\mu}_{11}$	$\hat{\mu}_{12}$	$\hat{\mu}_{13}$	$\hat{\mu}_{14}$
		B <sub>2</sub>	C <sub>2</sub>	2	$y_{121}$ ⋮ $y_{N_21}$	$y_{122}$ ⋮ $y_{N_22}$	$y_{123}$ ⋮ $y_{N_23}$	$y_{124}$ ⋮ $y_{N_24}$
					$\hat{\mu}_{21}$	$\hat{\mu}_{22}$	$\hat{\mu}_{23}$	$\hat{\mu}_{24}$
	A <sub>2</sub>	B <sub>1</sub>	C <sub>2</sub>	3	$y_{131}$ ⋮ $y_{N_31}$	$y_{132}$ ⋮ $y_{N_32}$	$y_{133}$ ⋮ $y_{N_33}$	$y_{134}$ ⋮ $y_{N_34}$
					$\hat{\mu}_{31}$	$\hat{\mu}_{32}$	$\hat{\mu}_{33}$	$\hat{\mu}_{34}$
		B <sub>2</sub>	C <sub>1</sub>	4	$y_{141}$ ⋮ $y_{N_41}$	$y_{142}$ ⋮ $y_{N_42}$	$y_{143}$ ⋮ $y_{N_43}$	$y_{144}$ ⋮ $y_{N_44}$
					$\hat{\mu}_{41}$	$\hat{\mu}_{42}$	$\hat{\mu}_{43}$	$\hat{\mu}_{44}$

Bei einem dreifaktoriellen Design beispielsweise, ordnet man die ersten beiden Faktoren A und B wie bei einem kreuzfaktoriellen Design an, und dann den dritten Faktor C so, daß jede seiner Stufen jeweils nur auf einer Stufe der anderen beiden Faktoren vorkommt (siehe Tabelle 3.5). Bei zweistufigen Faktoren führt das dazu, daß man statt  $2 \cdot 2 \cdot 2 = 8$  Zellen im kreuzfaktoriellen Design im lateinischen Quadrat nur  $2 \cdot 2 = 4$  Zellen mit Beobachtungseinheiten zu füllen hat. Der Nachteil besteht darin, daß keine Interaktionshypothesen geprüft werden können.

Zu prüfen ist zunächst die Hypothese, daß überhaupt keine Unterschiede in den Erwartungswertvektoren der P abhängigen Variablen zwischen den vier Zellen bestehen. In der Form  $C_B A = \Delta$  wählen wir dafür z.B.  $A_{A,B,C} = I$ ,  $\Delta_{A,B,C} = O$  und

$$(3.8.1) \quad C_{A,B,C} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}.$$

Die zweite Hypothese, daß keine Unterschiede in den Erwartungswertvektoren der P abhängigen Variablen zwischen den Stufen  $A_1$  und  $A_2$  bestehen, formulieren wir durch  $A_A = I$ ,  $\Delta_A = 0$  und

$$(3.8.2) \quad C_A = [1 \quad 1 \quad -1 \quad -1],$$

die entsprechende Hypothese für  $B_1$  und  $B_2$  durch  $A_B = I$ ,  $\Delta_B = 0$  und

$$(3.8.3) \quad C_B = [1 \quad -1 \quad 1 \quad -1],$$

sowie für  $C_1$  und  $C_2$  durch  $A_C = I$ ,  $\Delta_C = 0$  und

$$(3.8.4) \quad C_C = [1 \quad -1 \quad -1 \quad 1].$$

### 3.9 Lateinisches Quadrat über den abhängigen Variablen

Auch die abhängigen Variablen können im „lateinischen Quadrat“ strukturiert werden, sofern sie in mehrere Faktoren U, V und W mit gleicher Stufenzahl gegliedert werden können (siehe Tabelle 3.6). Die Ökonomie des Verfahrens besteht darin, daß wir bei z.B. zweistufigen Faktoren nur  $P = 2 \cdot 2 = 4$  abhängige Variablen erheben müssen anstelle von  $P = 2 \cdot 2 \cdot 2 = 8$  abhängigen Variablen im vergleichbaren kreuzfaktoriellen Design. Allerdings können auch hier (wie in Abschnitt 3.8) keine Interaktionshypothesen über die abhängigen Variablen selbst geprüft werden, wohl aber über Wechselwirkungen zwischen den unabhängigen und den abhängigen Variablen (vgl. Abschnitt 3.5).

Tabelle 3.6.: Faktoren U, V und W im „lateinischen Quadrat“ als Gliederung der P abhängigen Variablen  $y_p$ , welche insgesamt an  $N = N_1 + \dots + N_4$  Beobachtungseinheiten erhoben wurden. Beispiel für Faktor U: Meßmethode 1 ( $U_1$ ) Meßmethode 2 ( $U_2$ ); Faktor V: vormittags ( $V_1$ ), nachmittags ( $V_2$ ); Faktor W: Geschwindigkeit ( $W_1$ ), Genauigkeit ( $W_2$ ). Das Datenschema zeigt die  $P = 4$  abhängigen Variablen  $y_p$  pro Beobachtungseinheit und deren bedingte Erwartungswerte  $\hat{\mu}_{qp}$ .

### Gliederung der abhängigen Variablen

		$U_1$	$U_2$	Faktor U
		$V_1$	$V_2$	Faktor V
		$W_1$	$W_2$	Faktor W
Zellennummer q	1	2	3	Variablennummer p
1	$y_{111}$ ⋮ $y_{N_111}$	$y_{112}$ ⋮ $y_{N_112}$	$y_{113}$ ⋮ $y_{N_113}$	$y_{114}$ ⋮ $y_{N_114}$
	$\hat{\mu}_{11}$	$\hat{\mu}_{12}$	$\hat{\mu}_{13}$	$\hat{\mu}_{14}$
	$y_{121}$ ⋮ $y_{N_211}$	$y_{122}$ ⋮ $y_{N_212}$	$y_{123}$ ⋮ $y_{N_213}$	$y_{124}$ ⋮ $y_{N_214}$
	$\hat{\mu}_{21}$	$\hat{\mu}_{22}$	$\hat{\mu}_{23}$	$\hat{\mu}_{24}$
3	$y_{131}$ ⋮ $y_{N_311}$	$y_{132}$ ⋮ $y_{N_312}$	$y_{133}$ ⋮ $y_{N_313}$	$y_{134}$ ⋮ $y_{N_314}$
	$\hat{\mu}_{31}$	$\hat{\mu}_{32}$	$\hat{\mu}_{33}$	$\hat{\mu}_{34}$
	$y_{141}$ ⋮ $y_{N_411}$	$y_{142}$ ⋮ $y_{N_412}$	$y_{143}$ ⋮ $y_{N_413}$	$y_{144}$ ⋮ $y_{N_414}$
	$\hat{\mu}_{41}$	$\hat{\mu}_{42}$	$\hat{\mu}_{43}$	$\hat{\mu}_{44}$

Wir prüfen zunächst die globale Hypothese, daß in jeder Zelle die bedingten Erwartungswerte der P abhängigen Variablen gleich sind. In der Form  $\mathbf{C} \mathbf{B} \mathbf{A} = \Delta$  wählen wir z.B.  $\mathbf{C}_{U,V,W} = I$ ,  $\Delta_{U,V,W} = O$  und

$$(3.9.1) \quad \mathbf{A}_{U,V,W} = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{pmatrix}.$$

Als zweite Hypothese prüfen wir die Gleichheit der bedingten Erwartungswerte der Stufen  $U_1$  und  $U_2$  mit

$$(3.9.2) \quad \mathbf{A}_U = \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}$$

bei  $\mathbf{C}_U = I$  und  $\Delta_U = 0$ . Die entsprechenden Hypothesen für  $V_1$  und  $V_2$  bzw.  $W_1$  und  $W_2$  prüfen wir mit

$$(3.9.3) \quad \mathbf{A}_V = \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}$$

bzw.

$$(3.9.4) \quad \mathbf{A}_W = \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}$$

bei  $\mathbf{C}_V = \mathbf{C}_W = I$  und  $\Delta_V = \Delta_W = 0$ .

### 3.10 Zusammenfassende Bemerkungen

In diesem dritten Abschnitt wurde die Hypothesenformulierung in verschiedenen Designs für das Zellenmittelwertemodell bei gleichen Zellenfrequenzen behandelt. Dabei wurden exemplarisch drei Prinzipien zur Strukturierung von Variablen dargestellt, nämlich die kreuzfaktorielle Strukturierung, die hierarchische, und die nach dem „lateinischen Quadrat“. Es wurde gezeigt, wie diese Strukturierungsgesichtspunkte sowohl für die unabhängigen als auch für die abhängigen Variablen verwendet werden können. Dabei wurde z.B. eine wiederholt gemessene Variable formal als weitere abhängige Variable betrachtet, was dann angezeigt ist, wenn die Residuen der wiederholt gemessenen Varia-

blen zwischen den Meßpunkten nicht als unabhängig angenommen werden können (siehe Gleichung 2.3.11).

*Weiterführende Literatur:* Zur Klassifikation von Designs s. z.B. Bredenkamp (1980), Campbell und Stanley (1966) oder Kirk (1972, 1978). Winer (1971) gibt für viele Designs Rechenformeln an. Andere Arten von Vergleichen oder Kontrasten stellen fast alle der im Abschnitt 2 angegebenen Monographien dar, genauso wie auch andere Arten von Modellen, die anstelle des Zellenmittelwertemodells verwendet werden können.

## 4. Parameterschätzung

### 4.1 Einleitung

Die Parameter in der  $Q \times P$ -Matrix  $B$  eines multivariaten linearen Stichprobenmodells, welches die Gleichungen

$$(4.1.1) \quad E(Y) = X B$$

$$(4.1.2) \quad C(\boldsymbol{\varepsilon}_n, \boldsymbol{\varepsilon}_{n^*}) = C(y_n, y_{n^*}) = \begin{cases} \Sigma, & \text{für } n = n^* \\ O, & \text{für } n \neq n^* \end{cases} \quad n, n^* \in \{1, 2, \dots, N\}$$

erfüllt, wobei  $\boldsymbol{\varepsilon}_n := y_n - x_n B$ , sowie die Parameter in der  $P \times P$ -Kovarianzmatrix  $\Sigma$  können prinzipiell auf verschiedene Arten geschätzt werden. Keine der verschiedenen Schätzmethoden ist für alle Situationen optimal, und es würde den Rahmen des Artikels sprengen, wenn wir versuchen würden, alle Schätzmethoden darzustellen und kritisch zu bewerten. Statt dessen beschränken wir uns auf die Darstellung der Schätzung nach dem *Kriterium der kleinsten Quadrate ohne* (Abschnitt 4.2) *und mit Nebenbedingungen* (Abschnitt 4.3) sowie auf die *Maximum-Likelihood-Schätzung* unter der Annahme, daß die  $\boldsymbol{\varepsilon}_n$  jeweils  $P$ -variat normalverteilt sind, mit gleicher Kovarianzmatrix  $\Sigma_n = \Sigma$  für alle  $n \in \{1, 2, \dots, N\}$  (Abschnitt 4.4). Die Erwartungswerte und die Kovarianzmatrix der Schätzer von  $B$  und  $\Psi := CBA$  sind Gegenstand des Abschnitts 4.5, ebenso wie die Schätzung der Kovarianzmatrix  $\Sigma$ . Abschließend geben wir einige Hinweise auf weiterführende Literatur zu anderen Schätzmethoden und einigen Problemen, die in Zusammenhang mit der Parameterschätzung stehen. Die Schätztheorie, die in diesem Abschnitt dargestellt wird, ist nicht auf das Zellenmittelwertemodell beschränkt. Sie ist für alle Modelle anwendbar, welche die Gleichungen 4.1.1 und 4.1.2 erfüllen. Darunter fallen auch *Regressionsmodelle mit „nichtstochastischen Regressoren“*.

## 4.2 Kriterium der kleinsten Quadrate

Die Standardmethode, die Parametermatrix  $B$  zu schätzen, ist, den Schätzer  $\hat{B}$  so zu wählen, daß die *Spur der Fehlerquadratsummenmatrix*

$$(4.2.1) \quad f(\hat{B}) = \text{Spur} [(\mathbf{Y} - \mathbf{X} \hat{B})' (\mathbf{Y} - \mathbf{X} \hat{B})] = \\ = \text{Spur} [\mathbf{Y}'\mathbf{Y} - 2 \hat{B}'\mathbf{X}'\mathbf{Y} + \hat{B}'\mathbf{X}'\mathbf{X} \hat{B}]$$

ein Minimum wird (vgl. z.B. Timm, 1975, S. 185). Dies erreichen wir durch Differenzieren der Funktion  $f(B)$  nach  $B$  (s. Bock, 1975, S. 68-69)

$$(4.2.2) \quad \frac{\partial f(\hat{B})}{\partial \hat{B}} = -2 \mathbf{X}'\mathbf{Y} + 2 \mathbf{X}'\mathbf{X} \hat{B}$$

(vgl. z.B. Timm, 1975, S. 100), und Nullsetzen der resultierenden Gleichung:

$$(4.2.3) \quad \frac{\partial f(\hat{B})}{\partial \hat{B}} = \mathbf{O}.$$

Einen Schätzer  $\hat{B}$ , der diese Bedingung erfüllt, notieren wir mit  $\hat{B}$ . Die Bildung der zweiten Ableitung zeigt, daß die Funktion  $f(B)$  an der Stelle  $\hat{B} = \hat{B}$  tatsächlich ein Minimum aufweist.

Aus den Gleichungen 4.2.2 und 4.2.3 erhalten wir dann die sogenannte Normalgleichung

$$(4.2.4) \quad \mathbf{X}'\mathbf{X} \hat{B} = \mathbf{X}'\mathbf{Y},$$

die unter der Annahme der Nonsingularität von  $\mathbf{X}'\mathbf{X}$  zum Schätzer

$$(4.2.5) \quad \hat{B} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$$

führt. Man beachte die analogen Strukturen dieser Gleichung zu der Identifikationsgleichung 2.2.15.

Ist  $\mathbf{X}'\mathbf{X}$  jedoch singulär, so kann man dennoch mit der allgemeinen Inversen (vgl. z.B. Rao & Mitra, 1971)  $(\mathbf{X}'\mathbf{X})^{-}$ , welche die Eigenschaft  $\mathbf{X}'\mathbf{X} (\mathbf{X}'\mathbf{X})^{-} \mathbf{X}'\mathbf{X} = \mathbf{X}'\mathbf{X}$  erfüllt, eine Lösung für  $B$  angeben, nämlich

$$(4.2.6) \quad \hat{B} = (\mathbf{X}'\mathbf{X})^{-} \mathbf{X}'\mathbf{Y} + (\mathbf{I} - \mathbf{G}) \mathbf{Z}$$

wobei  $\mathbf{G} = (\mathbf{X}'\mathbf{X})^{-} \mathbf{X}'\mathbf{X}$  und  $\mathbf{Z}$  eine beliebige  $Q \times P$ -Matrix ist (vgl. z.B. Timm, 1975, S. 185). Dieser Schätzer ist im allgemeinen jedoch nicht eindeutig, da man  $\mathbf{Z}$  beliebig wählen kann, d.h. es gibt unendlich viele dieser Schätzer. Erst wenn  $\mathbf{X}'\mathbf{X}$  nonsingulär ist, d.h. wenn alle Spaltenvektoren von  $\mathbf{X}$  linear unab-

hängig sind und damit der Rang R von X gleich der Anzahl Q der Spalten von X ist, wird  $\hat{\mathbf{B}}$  eindeutig, und Gleichung 4.2.5 ist mit Gleichung 4.2.6 äquivalent. Eine andere Möglichkeit, einen eindeutigen Schätzer zu erhalten, besteht in der Aufstellung von Nebenbedingungen (Restriktionen) über B.

### 4.3 Kriterium der kleinsten Quadrate unter Nebenbedingungen

In vielen Fällen will man die Parametermatrix  $\mathbf{B}$  unter bestimmten Nebenbedingungen schätzen, die man z.B. in der Form  $\mathbf{H} \mathbf{B} = \Theta$  aufschreiben kann.

Dabei setzen wir voraus, daß die Anzahl der Zeilen der Matrix H gleich ihrem Rang ist. Solche Nebenbedingungen können zum einen Restriktionen sein, die erst die Schätzbarkeit von B ermöglichen, wie z.B.  $\sum_i \alpha_i = 0$  in der klassischen Varianzanalyse. Zum anderen kann es sich aber auch um solche Restriktionen handeln, die aus inhaltlich theoretischen Gründen gesetzt werden, wie etwa, daß alle Interaktionsparameter gleich Null sind, was auch ermöglicht, daß Interaktions- und Fehlervarianz zu einer neuen Fehlervarianzschätzung „gepoolt“ werden können.

Zur Minimierung der Funktion  $f(B)$  (siehe Gleichung 4.2.1) unter der Nebenbedingung  $\mathbf{H} \mathbf{B} = \Theta$  führt man eine Matrix  $\Lambda$  von Lagrange-Multiplikatoren ein, so daß die Funktion

$$(4.3.1) \quad g(\hat{\mathbf{B}}, \hat{\Lambda}) = \text{Spur} [(\mathbf{Y} - \mathbf{X} \hat{\mathbf{B}})' (\mathbf{Y} - \mathbf{X} \hat{\mathbf{B}})] + \\ + 2 \text{Spur} [\hat{\Lambda}' (\mathbf{H} \hat{\mathbf{B}} - \Theta)]$$

zu minimieren ist (vgl. z.B. Timm, 1975, S. 189).

Wir bilden zunächst die partiellen Ableitungen (s. Bock, 1975, S. 68-69)

$$(4.3.2) \quad \frac{\partial [g(\hat{\mathbf{B}}, \hat{\Lambda})]}{\partial (\hat{\mathbf{B}})} = -2 \mathbf{X}' \mathbf{Y} + 2 \mathbf{X}' \mathbf{X} \hat{\mathbf{B}} + 2 \mathbf{H}' \hat{\Lambda}$$

und

$$(4.3.3) \quad \frac{\partial [g(\hat{\mathbf{B}}, \hat{\Lambda})]}{\partial (\hat{\Lambda})} = \mathbf{H} \hat{\mathbf{B}} - \Theta.$$

Schätzer  $\hat{\mathbf{B}}$  und  $\hat{\Lambda}$ , welche die Bedingungen

$$(4.3.4) \quad \frac{\partial [g(\hat{\mathbf{B}}, \hat{\Lambda})]}{\partial (\hat{\mathbf{B}})} = \mathbf{O}$$

und

$$(4.3.5) \quad \frac{\partial g(\hat{\mathbf{B}}, \hat{\mathbf{\Lambda}})}{\partial (\hat{\mathbf{\Lambda}})} = \mathbf{O}$$

erfüllen, bezeichnen wir mit  $\hat{\mathbf{B}}_H$  bzw.  $\hat{\mathbf{\Lambda}}_H$ . Die zweite Ableitung zeigt, daß die Funktion  $g(\mathbf{B}, \mathbf{\Lambda})$  an der Stelle  $\hat{\mathbf{B}} = \hat{\mathbf{B}}_H$  bzw.  $\mathbf{\Lambda} = \hat{\mathbf{\Lambda}}_H$  tatsächlich ein Minimum aufweist.

Aus den Gleichungen 4.3.2 bis 4.3.5 erhalten wir

$$(4.3.6) \quad \mathbf{X}'\mathbf{X} \hat{\mathbf{B}}_H = \mathbf{X}'\mathbf{Y} - \mathbf{H}'\hat{\mathbf{\Lambda}}_H$$

und

$$(4.3.7) \quad \mathbf{H} \hat{\mathbf{B}}_H = \mathbf{\Theta}.$$

Man benutzt nun die für ein konsistentes Gleichungssystem  $\mathbf{AB} = \mathbf{C}$  (mit der Matrix  $\mathbf{B}$  von Unbekannten) gültige allgemeine Lösung

$$(4.3.8) \quad \mathbf{B} = \mathbf{A}^{-1}\mathbf{C} + (\mathbf{I} - \mathbf{A}^{-1}\mathbf{A})\mathbf{Z},$$

wobei  $\mathbf{Z}$  eine beliebige Matrix passenden Typs ist (siehe z.B. Timm 1975, S. 55). Auf die Gleichung 4.3.6 angewandt erhält man daraus für  $\hat{\mathbf{B}}_H$  die Lösung

$$(4.3.9) \quad \hat{\mathbf{B}}_H = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}'\hat{\mathbf{\Lambda}}_H + [\mathbf{I} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}] \mathbf{Z}.$$

Unter Verwendung von Gleichung 4.2.6 folgt dann

$$(4.3.10) \quad \hat{\mathbf{B}}_H = \hat{\mathbf{B}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}'\hat{\mathbf{\Lambda}}_H.$$

Vormultiplikation mit  $[\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}']^{-1}\mathbf{H}$  und Einsetzen von Gleichung 4.3.7 ergibt

$$(4.3.11) \quad \hat{\mathbf{\Lambda}}_H = [\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}']^{-1}(\mathbf{H}\hat{\mathbf{B}} - \mathbf{\Theta}).$$

Aus den Gleichungen 4.3.10 und 4.3.11 erhalten wir dann

$$(4.3.12) \quad \hat{\mathbf{B}}_H = \hat{\mathbf{B}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}'[\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}']^{-1}(\mathbf{H}\hat{\mathbf{B}} - \mathbf{\Theta}).$$

Hat Gleichung 4.3.12 eine eindeutige Lösung, so ist  $\hat{\mathbf{B}}_H$  der Kleinsten-Quadrat-Schätzer für die Parametermatrix  $\mathbf{B}$  unter der Nebenbedingung  $\mathbf{HB} = \mathbf{\Theta}$  (vgl. z.B. Searle, 1971, S. 204ff. oder Timm, 1975, S. 178).

## 4.4 Maximum-Likelihood-Kriterium

Bei den Schätzern nach dem Kriterium der kleinsten Quadrate müssen außer der Gleichheit der Kovarianzmatrizen (siehe Gleichung 4.1.2) keine Verteilungsannahmen getroffen werden, es sei denn, zum Nachweis einiger Optimalitätseigenschaften. Ist man jedoch zum Beispiel in der Lage anzunehmen, daß die Dichtefunktion der N Vektoren  $\boldsymbol{\epsilon}_n$  der Residualvariablen eine P-variate Normalverteilung mit den PxP-Kovarianzmatrizen  $\boldsymbol{\Sigma}_n = \boldsymbol{\Sigma}$  ist, daß die  $\boldsymbol{\epsilon}_n$  stochastisch unabhängig sind, daß  $N \geq P+Q$  und daß der Rang R von X gleich Q, der Anzahl ihrer Spalten, ist, dann führt die Maximierung der Likelihoodfunktion

$$L(\hat{\mathbf{B}}) = 2 \pi^{-(1/2)PN} |\boldsymbol{\Sigma}|^{-(1/2)N} \exp \left[ -\frac{1}{2} \sum_{n=1}^N (\mathbf{y}_n - \mathbf{x}_n \hat{\mathbf{B}})' \boldsymbol{\Sigma}^{-1} (\mathbf{y}_n - \mathbf{x}_n \hat{\mathbf{B}}) \right]$$

bezüglich  $\hat{\mathbf{B}}$  zu dem Maximum-Likelihood-Schätzer

$$(4.4.1) \quad \hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$$

(vgl. z.B. Anderson, 1958, S. 181), der also unter den genannten Bedingungen mit dem Kleinsten-Quadrat-Schätzer identisch ist.

## 4.5 Erwartungswerte- und Kovarianzmatrix der Parametervektoren $\beta_p$ und $\psi_k$

Wir zeigen zunächst, daß der Schätzer

$$(4.5.1) \quad \hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$$

ein erwartungstreuer Schätzer von B ist:

$$(4.5.2) \quad E(\hat{\mathbf{B}}) = E [(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}] = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'E(\mathbf{Y}) = \\ = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{X} \mathbf{B} = \mathbf{B},$$

wobei wir lediglich  $E(Y) = X B$  (siehe Gleichung 4.1.1) verwendet haben. Entsprechend gilt für den Schätzer

$$(4.5.3) \quad \hat{\boldsymbol{\Psi}} = \mathbf{C}\hat{\mathbf{B}}\mathbf{A}$$

die Gleichung

$$(4.5.4) \quad E(\hat{\boldsymbol{\Psi}}) = E(\mathbf{C}\hat{\mathbf{B}}\mathbf{A}) = \mathbf{CBA}.$$

Die Kovarianzmatrix von einem Spaltenvektor  $\hat{\beta}_p$  der Parametermatrix  $\hat{B}$  ist

$$(4.5.5) \quad C(\hat{\beta}_p, \hat{\beta}_p) = C[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'y_p, (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y_p] = \\ = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'C(y_p, y_p) \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \\ = \sigma_p^2 (\mathbf{X}'\mathbf{X})^{-1},$$

wobei  $y_p$  die p-te Spalte der Matrix Y ist. Die Kovarianzmatrix  $C(y_p, y_p)$  ist nämlich die NxN-Diagonalmatrix  $\sigma_p^2 I$  und  $\sigma_p^2 = \sigma_{pp}$  die Varianz von  $y_p$ , der p-ten, Komponente des Vektors der abhängigen Variablen. Die Matrix  $C(\hat{\beta}_p, \hat{\beta}_p)$  wird z.B. für die Bestimmung von Konfidenzintervallen der  $\beta_{qp}$  benötigt (siehe Abschnitt 5.6).

Für die Kovarianzmatrix eines Mxl-Spaltenvektors  $\hat{\psi}_k = \mathbf{C}\hat{B}\mathbf{a}_k$  welche z.B. für die Angabe von Konfidenzintervallen bestimmter Linearkombinationen von B benötigt wird, gilt

$$(4.5.6) \quad C(\hat{\psi}_k, \hat{\psi}_k) = C[\mathbf{C}\hat{B}\mathbf{a}_k, \mathbf{C}\hat{B}\mathbf{a}_k] = \mathbf{C} C(\hat{B}\mathbf{a}_k, \hat{B}\mathbf{a}_k) \mathbf{C}' = \\ = \mathbf{C}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'[C(Y\mathbf{a}_k, Y\mathbf{a}_k)] \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{C}' = \\ = \mathbf{C}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'[V(y_n\mathbf{a}_k) I]\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{C}' = \\ = V(y_n\mathbf{a}_k) \mathbf{C}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{C}',$$

wobei  $V(y_n\mathbf{a}_k) I$  eine NxN-Diagonalmatrix und  $V(y_n\mathbf{a}_k)$  die Varianz von  $y_n\mathbf{a}_k$  ist, die wegen der Annahme 4.1.2 für alle N Zeilen von Y gleich ist. Für die N Varianzen  $V(y_n\mathbf{a}_k)$  gilt dabei

$$(4.5.7) \quad V(y_n\mathbf{a}_k) = C(y_n\mathbf{a}_k, y_n\mathbf{a}_k) = \mathbf{a}'_k C(y_n, y_n) \mathbf{a}_k = \mathbf{a}'_k \Sigma \mathbf{a}_k.$$

Die PxP-Kovarianzmatrix  $\Sigma$  kann durch die folgenden Schätzer erwartungstreu geschätzt werden:

$$(4.5.8) \quad \hat{\Sigma} = \frac{1}{N-R} (\mathbf{Y} - \mathbf{X} \hat{B})'(\mathbf{Y} - \mathbf{X} \hat{B}) = \\ = \frac{1}{N-R} \mathbf{Y}'[I - \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{Y}$$

(siehe z.B. Timm, 1975, S. 186).

## 4.6 Zusammenfassende Bemerkungen

In diesem Abschnitt wurden die Standardschätzmethoden für die Parameter des Modells  $E(Y) = X \beta$  dargestellt. Dabei ist insbesondere die in Abschnitt 4.3 angegebene Schätzmethode nach dem Kriterium der kleinsten Quadrate unter einer Nebenbedingung auch in solchen Fällen anwendbar, in denen die Matrix  $X$  keinen vollen Spaltenrang hat. Dies ist z.B. bei klassischen varianzanalytischen Modellen der Fall, bei denen man diesen „Mangel“ durch Nebenbedingungen wie z.B.  $\sum \alpha_j = 0$  „beheben“ muß. Ein Einblick in die Komplexität des Themas „Schätzmethoden“ soll andeutungsweise in den Bemerkungen zur weiterführenden Literatur vermittelt werden.

*Weiterführende Literatur:* Grundlage für eine kritische Bewertung verschiedener Schätzmethoden sind Gütekriterien für Schätzer, wie Erwartungstreue, Effizienz, Robustheit etc. Eine gut lesbare Einführung ist z.B. Burkholder (1978, S. 251ff.). Als anspruchsvolle und ausführliche Arbeit sei dazu Zacks (1971) genannt. Andrews et al. (1972) geben einen Überblick zum Thema „Robustheit“.

Andere Schätzmethoden als die hier dargestellten werden z.B. in Bandemer et al. (1977) dargestellt. Eine wichtige Alternative ist z.B. die *verallgemeinerte Methode der kleinsten Quadrate*, die dann angewandt werden kann, wenn die Kovarianzmatrizen  $C(y_n, y_n)$  nicht mehr für alle  $N$  Beobachtungseinheiten als gleich angenommen werden können (vgl. z.B. Searle, 1971, S. 87 oder Humak, 1977, S. 30). Eine weitere wichtige Schätzmethode ist die *Bayer-Schätzung*, die vor allem dann gewinnbringend verwendet werden kann, wenn Vorwissen über die Parameter besteht. Dazu verweisen wir auf Bandemer et al. (1977), Humak (1977), Leonhard (1975) sowie Lindley und Smith (1972).

Bei der *Stein'schen Schätzung* besteht der Vorteil darin, daß extreme Werte („Ausreißer“) den Schätzer weniger stark beeinflussen als beim Kriterium der kleinsten Quadrate (vgl. Bandemer et al., 1977, S. 111), was man auch mit der  *$L_p$ -Approximation* (vgl. z.B. Späth, 1973, 1974) erreichen kann, bei der das Minimierungskriterium nicht unbedingt die Summe der Abweichungsquadrate ( $L_2$ ), sondern z.B. die der absoluten Beträge der Abweichungen ( $L_1$ ) ist. Schließlich sei noch auf Harter (1974-1975) hingewiesen, der eine ausführliche Bibliographie über das „Kriterium der kleinsten Quadrate und einige Alternativen“ (Übersetzung durch die Autoren) liefert.

## 5. Hypothesenbewertung

### 5.1 Einleitung

In Abschnitt 3 haben wir ausführlich für verschiedene Designs gezeigt, wie man Hypothesen über die Parameter in der Matrix  $B$  des multivariaten linearen Stichprobenmodells

$$(5.1.1) \quad E(\mathbf{Y}) = \mathbf{XB}$$

formulieren kann, und zwar in der Form der allgemeinen multivariaten linearen Hypothese

$$(5.1.2) \quad H_0: \Psi := \mathbf{CBA} = \Delta$$

mit der  $M \times Q$ -Matrix  $\mathbf{C}$ , der  $Q \times P$ -Matrix  $\mathbf{B}$ , der  $P \times K$ -Matrix  $\mathbf{A}$  und der  $M \times K$ -Matrix  $\mathbf{A}$ , wobei  $C$  den Rang  $M \leq R$  und  $A$  den Rang  $K \leq P \leq R$  hat und  $R$  der Rang der  $N \times Q$ -Matrix  $X$  ist. Dort hatten wir immer vorausgesetzt, daß die Parametermatrix  $\mathbf{B}$  die bedingten Erwartungswerte der  $P$  abhängigen Variablen in den  $Q$  Zellen enthält (siehe Abschnitt 3.2). Diese Voraussetzung brauchen wir hier nicht zu machen, d.h. die in diesem Abschnitt dargestellten Methoden zur Bewertung der Hypothese 5.1.2 gelten für alle Modelle, welche die Gleichung 5.1.1 und

$$(5.1.3) \quad C(y_n, y_{n^*}) = C(\boldsymbol{\varepsilon}_n, \boldsymbol{\varepsilon}_{n^*}) = \begin{cases} \Sigma, & \text{für } n = n^* \\ & n, n^* \in \{1, 2, \dots, N\}, \\ \mathbf{O}, & \text{für } n \neq n^* \end{cases}$$

erfüllen, wobei  $\boldsymbol{\varepsilon}_n = y_n - \mathbf{x}_n \mathbf{B}$ , und bei denen  $\Psi := \mathbf{CBA}$  schätzbar ist (vgl. z.B. Timm, 1975, S. 173). Auch der Abschnitt über Parameterschätzung war nicht auf das Zellenmittelwertemodell beschränkt.

Für den multivariaten Fall wurden eine Reihe von Testkriterien vorgeschlagen, von denen wir die vier wichtigsten darstellen. Dabei folgen wir der Auswahl von Olson (1976), der auch eine ausführliche Diskussion über die Vor- und Nachteile der Kriterien in verschiedenen Situationen liefert (vgl. auch Stevens, 1979). Bei den Verteilungen aller vier Testkriterien wird die  $P$ -variate Normalverteilung der Zeilenvektoren  $y_n$ ,  $n \in \{1, 2, \dots, N\}$ , und deren Unabhängigkeit voneinander (siehe Gleichung 5.1.3) vorausgesetzt, und bei allen vier Testkriterien spielen die beiden Matrizen  $Q_e$  und  $Q_h$  eine wesentliche Rolle, die wir im folgenden Absatz behandeln.

Die Matrix

$$(5.1.4) \quad Q_e = (\mathbf{YA} - \mathbf{X}\hat{\mathbf{B}}\mathbf{A})'(\mathbf{YA} - \mathbf{X}\hat{\mathbf{B}}\mathbf{A}) = \\ = \mathbf{A}'\mathbf{Y}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{YA},$$

die wir *Fehlerquadratsummenmatrix* nennen, folgt unter den oben angegebenen Voraussetzungen der zentralen  $K$ -variaten Wishart-Verteilung  $\mathbf{W}_X(N-R, \mathbf{A}\Sigma\mathbf{A}')$  (siehe z.B. Johnson & Kotz, 1972, S. 158).

Die Matrix

$$(5.1.5) \quad Q_h = (\mathbf{C}\hat{\mathbf{B}}\mathbf{A} - \Delta)' [\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1} (\mathbf{C}\hat{\mathbf{B}}\mathbf{A} - \Delta),$$

die wir *Hypothesenquadratsummenmatrix* nennen, folgt ebenfalls der zentralen K-variaten Wishart-Verteilung  $W_K(M, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$ , falls  $\mathbf{\Psi} = \mathbf{CBA} = \mathbf{\Delta}$ , d.h. wenn die Nullhypothese richtig ist. Zur Ableitung dieser Gleichung siehe z.B. Timm (1975, S. 183).

## 5.2 Wilks' Lambda-Kriterium

Unter den in Abschnitt 5.1 angegebenen Voraussetzungen folgt das *Lambda-Kriterium* von Wilks (1932)

$$(5.2.1) \quad \Lambda := \frac{|\mathbf{Q}_e|}{|\mathbf{Q}_e + \mathbf{Q}_h|}$$

der zentralen Wilks' Lambda-Verteilung  $\Lambda(K, M, N-R)$ , falls  $\mathbf{CBA} = \mathbf{\Delta}$ . Zum Beweis dieser Behauptung verweisen wir auf Timm (1975, S. 189ff), bei dem man auch eine Tabelle für die Verteilung von  $\Lambda$  findet.

Eine häufig anstelle von  $\Lambda$  verwendete Statistik, die auf Rao zurückgeht (vgl. Rao, 1973, S. 556), ist

$$(5.2.2) \quad F = \frac{1 - \Lambda^{1/a}}{\Lambda^{1/a}} \cdot \frac{ab - 2c}{K \cdot M},$$

die unter den genannten Bedingungen approximativ der zentralen F-Verteilung folgt mit den Freiheitsgraden  $K \cdot M$  und  $ab - 2c$ , wobei

$$(5.2.3) \quad a = \begin{cases} 1, & \text{falls } K \cdot M = 2 \\ \sqrt{\frac{K^2 \cdot M^2 - 4}{K^2 + M^2 - 5}}, & \text{andernfalls,} \end{cases}$$

sowie

$$(5.2.4) \quad b = N - R + M - \frac{K + M + 1}{2}$$

und

$$(5.2.5) \quad c = \frac{K \cdot M - 2}{4}$$

sind. Wenn  $(ab - 2c)$  keine ganze Zahl ist, kann man den ganzzahligen Anteil verwenden, um „konservativer“ zu testen.

Über die Berechnung von  $\Lambda$  als Quotient von verallgemeinerten Varianzen, wesentliche Eigenschaften von  $\Lambda$ , sowie eine approximativ X-verteilte Prüfstatistik nach Bartlett (1947) berichtet Moosbrugger (1978, S. 114-124).

### 5.3 Roy's Eigenwert-Kriterium

Roy (z.B. 1957) hat den größten Eigenwert der Matrix  $(Q_e + Q_h)^{-1} Q_h$ , nämlich den Wert  $\vartheta_{\max}$ , welcher die Determinantengleichung

$$(5.3.1) \quad |(Q_e + Q_h)^{-1} Q_h - \vartheta I| = 0$$

erfüllt (vgl. Moosbrugger, 1978, S. 45f.), als Testkriterium für die Nullhypothese

$$(5.3.2) \quad H_0: \bigcap_v (\mathbf{CBA}v = \Delta v)$$

vorgeschlagen, die nur dann wahr ist, wenn für jeden beliebigen K-dimensionalen Vektor  $v$  die Gleichung  $\mathbf{CBA}v = \Delta v$  gilt (vgl. z.B. Timm, 1975, S. 192).

Die Verteilung dieses Testkriteriums  $\vartheta_{\max}$  wurde von Heck (1960) und Pillai (1960) tabelliert mit den Argumenten

$$(5.3.3) \quad I = \min(K, M),$$

$$(5.3.4) \quad a = \frac{|K-M| - 1}{2}$$

und

$$(5.3.5) \quad b = \frac{N-R-K-1}{2}.$$

Dabei benutzt Pillai die Buchstaben s für I, m für a und n für b. Tabellen für dieses Testkriterium findet man aber auch bei Morrison (1967) und Timm (1975).

### 5.4 Hotelling-Lawley Spur Kriterium

Ein anderes Testkriterium, das auf Lawley (1938) und Hotelling (1951) zurückgeht, ist

$$(5.4.1) \quad U = Spur [Q_h Q_e^{-1}] = \sum_{i=1}^I \lambda_i,$$

wobei  $I = \min(K, M)$  und  $\lambda_i$  der i-te Eigenwert der Matrix  $Q_h Q_e^{-1}$  ist. Pillai (1960) hat die zentrale Verteilung dieser Statistik abgeleitet und tabelliert, ebenfalls mit den Argumenten I, a und b (siehe Gleichungen 5.3.3 bis 5.3.5), wobei Pillai auch hier die Buchstaben s für I, m für a und n für B benutzt. Auch bei Timm (1975) findet man eine Tabelle für diese Statistik.

## 5.5 Pillai-Bartlett Spur Kriterium

Ein weiteres Testkriterium, das *Pillai-Bartlett-Spur-Kriterium*, ist definiert als

$$(5.5.1) \quad V = I - \text{Spur} [\mathbf{Q}_e (\mathbf{Q}_h + \mathbf{Q}_e)^{-1}] = \sum_{i=1}^I (1 - \lambda_i),$$

wobei  $I = \min(K, M)$  und  $\lambda_i$  der  $i$ -te Eigenwert der Matrix  $\mathbf{Q}_e (\mathbf{Q}_h + \mathbf{Q}_e)^{-1}$  ist. Eine ausführliche Tabelle dieser Statistik findet man in Mijares (1964). Weniger umfangreiche Tabellen sind in Pillai (1960), Pillai und Jayachandran (1967, 1970), Pillai und Mijares (1959) sowie Timm (1975) abgedruckt.

Pillai (1960, S. 19) hat außerdem eine F-Approximation von  $V$  angegeben, nämlich

$$(5.5.2) \quad F = \frac{(N-R-K+I) \cdot V}{T \cdot (I-V)},$$

wobei  $T = \max(K, M)$ . Diese Statistik ist annähernd zentral F-verteilt mit  $J \cdot T$  und  $I \cdot (N-R-K+I)$  Freiheitsgraden.

Nach Olson (1976, S. 584) ist diese Approximation hinreichend genau und nur bei kleinen Freiheitsgraden leicht konservativ.

## 5.6 Einfache Konfidenzintervalle

In diesem Abschnitt werden einfache Konfidenzintervalle behandelt, im Unterschied zu simultanen, über welche z.B. Gabriel (1968, 1969 a, b) informiert.

### 5.6.1 Konfidenzintervall für einen Parameter

Aus den in Abschnitt 5.1 aufgeführten Annahmen folgt, daß

$$(5.6.1) \quad \frac{\hat{\beta}_{qp} - \beta_{qp}}{\sqrt{V(\hat{\beta}_{qp})}} \sim N(0,1)$$

mit Erwartungswert Null und Varianz Eins normalverteilt ist, wobei die Varianz  $V(\hat{\beta}_{qp})$  das  $q$ -te Element der Hauptdiagonalen der Kovarianzmatrix

$$(5.6.2) \quad C(\hat{\beta}_p, \hat{\beta}_p) = \sigma_p^2 (\mathbf{X}' \mathbf{X})^{-1}$$

(siehe Gleichung 4.5.5) ist (vgl. z.B. Searle, 1971, S. 107). Außerdem folgt aus diesen Annahmen, daß

$$(5.6.3) \quad \frac{\hat{\beta}_{qp} - \beta_{qp}}{\sqrt{V(\hat{\beta}_{qp})}} \sim t(N-R)$$

t-verteilt ist mit N-R Freiheitsgraden, wobei R der Rang der Matrix X ist und  $\hat{V}(\hat{\beta}_{qp})$  der Schätzer von  $V(\hat{\beta}_{qp})$ . Den Schätzer  $\hat{V}(\hat{\beta}_{qp})$  erhalten wir als q-tes Element der Hauptdiagonalen von

$$(5.6.4) \quad \hat{C}(\hat{\beta}_p, \hat{\beta}_p) = \hat{\sigma}_p^2 (\mathbf{X}'\mathbf{X})^{-1},$$

wobei  $\hat{\sigma}_p^2$  das p-te diagonale Element der Matrix  $\hat{\Sigma}$  (siehe Gleichung 4.5.8) ist.

Ein Konfidenzintervall für  $\beta_{qp}$  konstruieren wir auf folgende Weise: Es sei  $\pm t(N-R, \alpha/2)$  der untere bzw. der obere Punkt der t-Verteilung, für welche

$$(5.6.5) \quad P[-t(N-R, \alpha/2) \leq t \leq t(N-R, \alpha/2)] = 1 - \alpha$$

die Wahrscheinlichkeit ist, daß ein t-verteilter Wert in dem angegebenen Intervall liegt. Wendet man dies auf 5.6.3 an, so erhält man

$$(5.6.6) \quad P[-t(N-R, \alpha/2) \leq \frac{\hat{\beta}_{qp} - \beta_{qp}}{\sqrt{\hat{V}(\hat{\beta}_{qp})}} \leq t(N-R, \alpha/2)] = 1 - \alpha.$$

Daraus ergeben sich die Grenzen des symmetrischen Konfidenzintervalle

$$(5.6.7) \quad \hat{\beta}_{qp} - t(N-R, \alpha/2) \sqrt{\hat{V}(\hat{\beta}_{qp})} \leq \beta_{qp} \leq \hat{\beta}_{qp} + t(N-R, \alpha/2) \sqrt{\hat{V}(\hat{\beta}_{qp})}$$

für den unbekannten Parameter  $\beta_{qp}$ , in dem der Parameter  $\beta_{qp}$  mit Wahrscheinlichkeit  $1 - \alpha$  liegt.

Entsprechend kann auch für jede Linearkombination  $\mathbf{c}'\hat{\beta}_p$  der Koeffizienten eines Vektors  $\hat{\beta}_p$  (das ist eine Spalte der  $Q \times P$ -Parametermatrix B) ein Konfidenzintervall angegeben werden, dessen Grenzen wir durch

$$(5.6.8) \quad \mathbf{c}'\hat{\beta}_p \pm t(N-R, \alpha/2) \cdot \sqrt{\hat{V}(\mathbf{c}'\hat{\beta}_p)},$$

finden, wobei

$$(5.6.9) \quad \hat{V}(\mathbf{c}'\hat{\beta}_p) = \hat{\sigma}_p^2 \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1} \mathbf{c}$$

(siehe Gleichung 5.6.4).

Dabei kann  $\mathbf{c}'$  z.B. eine Zeile der Matrix C der allgemeinen multivariaten linearen Hypothese  $H_0 : \Psi = \mathbf{C}\mathbf{B}\mathbf{A} = \Delta$  sein, oder auch eine Zeile  $\mathbf{x}_n$  in der Matrix X des Modells  $\mathbf{E}(\mathbf{Y}) = \mathbf{X}\mathbf{B}$ . Im letzteren Fall ergibt 5.6.8 ein Konfidenzintervall für den Erwartungswert

$$(5.6.10) \quad E(\hat{y}_{np}) = E(\mathbf{x}_n \hat{\beta}_p)$$

des vorhergesagten Wertes  $\hat{y}_{np}$  der Beobachtungseinheit n auf der Variablen  $y_p$ .

### 5.6.2 Konfidenzintervall für Realisationen einer abhängigen Variablen $y_p$

Von größerem Interesse dürfte allerdings sein, in welchem Intervall eine Realisation der Beobachtungseinheit n in der abhängigen Variablen  $y_p$  liegt. Dieses Konfidenzintervall wird nach den folgenden Überlegungen konstruiert. Man geht aus von der Differenz

$$(5.6.11) \quad \hat{y}_{np} - y_{np} = \hat{y}_{np} - (\mathbf{x}_n \boldsymbol{\beta}_p + e_{np}),$$

welche die Varianz

$$(5.6.12) \quad V(\hat{y}_{np} - y_{np}) = V(\hat{y}_{np}) + V(e_{np})$$

hat, da  $\mathbf{x}_n \boldsymbol{\beta}_p$  eine Konstante und  $C(\hat{y}_{np}, e_{np}) = 0$  ist. Die stochastische Variable

$$(5.6.13) \quad \frac{\hat{y}_{np} - y_{np} - E(\hat{y}_{np} - y_{np})}{\sqrt{V(\hat{y}_{np} - y_{np})}} = \frac{\hat{y}_{np} - y_{np}}{\sqrt{V(\hat{y}_{np} - y_{np})}} \sim t(N-R)$$

ist wiederum t-verteilt mit N-R Freiheitsgraden, wobei N die Anzahl der Beobachtungseinheiten und R der Rang der Matrix X ist.

Analog zu dem oben für  $\beta_{qp}$  angestellten Überlegungen (siehe Gleichung 5.6.7) erhalten wir wieder

$$(5.6.14) \quad P\left[-t(N-R, \alpha/2) \leq \frac{\hat{y}_{np} - y_{np}}{\sqrt{V(\hat{y}_{np} - y_{np})}} \leq +t(N-R, \alpha/2)\right] = 1 - \alpha.$$

Daraus ergibt sich das symmetrische Konfidenzintervall

$$(5.6.15) \quad \begin{aligned} \hat{y}_{np} - t(N-R, \alpha/2) \cdot \sqrt{V(\hat{y}_{np} - y_{np})} &\leq y_{np} \leq \\ &\leq \hat{y}_{np} + t(N-R, \alpha/2) \cdot \sqrt{V(\hat{y}_{np} - y_{np})}, \end{aligned}$$

in dem mit Wahrscheinlichkeit  $1 - \alpha$  eine Realisation von  $y_{np}$  (ein noch unbeobachteter Wert der Beobachtungseinheit n auf der p-ten abhängigen Variablen) liegt.

Die dazu benötigte Varianzschätzung finden wir gemäß der Gleichung 5.6.12, wobei wir  $V(\hat{y}_{np})$  nach Gleichung 5.6.9 (mit  $c' = \mathbf{x}_n$ ) schätzen und  $\mathbf{V}(e_{np})$  durch  $\hat{\sigma}_p^2$ , dem p-ten Element der Diagonalmatrix  $\hat{\Sigma}$  (siehe Gleichung 4.5.8), als

$$(5.6.16) \quad \hat{V}(\hat{y}_{np} - y_{np}) = \hat{\sigma}_p^2 [1 + \mathbf{x}_n (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}'_n].$$

## 5.7 Praktische Signifikanz

Der in Abschnitt 2.2 (Gleichung 2.2.24) angegebene multivariate Determinationskoeffizient kann als praktisches Signifikanzmaß dienen und, falls  $C_1 = 0'$ ,  $1'A = 0'$  und  $\Delta = 0$  sind mit  $1$  bzw.  $1'$  als „passenden“ Einheitsvektoren, durch

$$(5.7.1) \quad R_H^2 = \frac{\text{Spur}[Q_H Q_T^{-1}]}{P}$$

geschätzt werden, mit der Hypothesenquadratsummenmatrix

$$(5.7.2) \quad Q_H = (\hat{C}B\mathbf{A} - \Delta)' [C(X'X)^{-1}C']^{-1} (\hat{C}B\mathbf{A} - \Delta)$$

und der Totalquadratsummenmatrix

$$(5.7.3) \quad Q_T = (YA - \bar{Y}A)' (YA - \bar{Y}A) = A'Y'YA - A'\bar{Y}'\bar{Y}A,$$

wobei

$$(5.7.4) \quad \bar{Y} = \frac{1}{N} \mathbf{1} \mathbf{1}' Y$$

eine NxP-Matrix ist, welche in jeder Zeile die Gesamtmittelwerte der P abhängigen Variablen über alle N Beobachtungseinheiten als Komponenten enthält.  $1$  bezeichnet dabei den Einheitsspaltenvektor, der aus N Einsen besteht.

$R_H^2 \cdot 100\%$  gibt bei konstant gehaltenem  $A = I$  an, wieviel Varianzprozente der abhängigen Variablen durch die Variationsquelle, auf welche sich die betreffende Hypothese bezieht, determiniert sind. Man beachte, daß man mit  $A \neq I$  in der multivariaten allgemeinen linearen Hypothese  $CBA - \Delta = 0$  einen neuen Vektor abhängiger Variablen

$$(5.7.5) \quad w = y A$$

bildet.

## 5.8 Zusammenfassende Bemerkungen

In diesem Abschnitt wurden vier Kriterien für multivariate Signifikanztests dargestellt, die alle invariant bezüglich linearer Transformationen der Ausgangsdaten sind. Im Gegensatz zum univariaten Fall, in dem der F-Test der gleichmäßig beste (uniformly most powerful) ist, gibt es im multivariaten Fall keine Teststatistik mit dieser Optimalitätseigenschaft. Statt dessen hängt die Teststärke der Kriterien davon ab, ob sich die Mittelwerteunterschiede zwischen den Gruppen auf eine Dimension konzentrieren („concentrated noncen-

trality structure“ (Olson, 1976, S. 581)) oder ob sie über mehrere Dimensionen der abhängigen Variablen verstreut („diffuse noncentrality structure“, ebenda) sind. Im ersten Fall ist Roys Kriterium 6 des größten Eigenwertes und im letzten Fall das Pillai-Bartlett-Spur-Kriterium V die Statistik mit der größten Teststärke. Für eine ausführlichere Diskussion verweisen wir auf Olson (1976) und Stevens (1979).

Ein Signifikanztest sollte immer von der Berechnung eines Maßes der Effektgröße begleitet sein. Bredenkamp (1970) diskutiert univariate „Maße der praktischen Signifikanz“. Für den multivariaten Fall geben wir den multivariaten Determinationskoeffizienten als praktisches Signifikanzmaß an, welcher auf den Überlegungen von Cramer und Nicewander (1979), Shaffer und Gillo (1974) sowie Stevens (1972) beruht. Die erörterten Konfidenzintervalle für die Modellparameter und für prädizierte Werte der abhängigen Variablen liefern nützliche Informationen zur Bemessung der statistischen Bedeutsamkeit von Zusammenhängen und Unterschieden, die sich in den Modellparametern  $\beta_{qp}$  niederschlagen.

*Weiterführende Literatur:* Die multivariaten Testkriterien werden in den Standardwerken zur multivariaten Analyse dargestellt (vgl. z.B. Bock, 1975, Finn, 1974, oder Timm, 1975). Eine wichtige Ergänzung zu Signifikanztests sind Verfahren multipler Vergleiche. Wir verweisen dazu auf die Monographie von Miller (1966) sowie auf Carmer und Swanson (1973), Einot und Gabriel (1975), Gabriel (1964, 1969a,b), Gernes (1971a,b, 1973), Petrinovich und Hardyck (1969), Ryan (1959) und Scheffé (1953). Ein nützlicher Sammelband für alle mit der Varianzanalyse zusammenhängenden Probleme ist Krishnaiah (1980).

## Literatur

- Ahrens & Läuter, 1981. Mehrdimensionale Varianzanalyse. Akademie-Verlag Berlin.
- Aitkin, M. A. 1969. Multiple comparisons in psychological experiments. British Journal of Mathematical and Statistical Psychology, 22, 193-198.
- Anderson, T. W. 1958. An introduction to multivariate statistical analysis. New York: Wiley.
- Andrews, D. F. et al. 1972. Robust estimates of location: Survey and advances: Princeton: University Press.
- Appelbaum, M. I. & Cramer, E. M. 1974. Some problems in the nonorthogonal analysis of variance. Psychological Bulletin, 81, 335-343.
- Bandemer, H. et al. 1977. Theorie und Anwendung der optimalen Versuchsplanung I. Berlin: Akademie-Verlag.
- Bartlett, M. S. 1947. Multivariate analysis. Journal of the Royal Statistical Society. Serie B, 9, 176-197.

- Bauer, H. 1974. Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie. Berlin: de Gruyter.
- Bock, R. D. 1975. Multivariate statistical methods in behavioral research. New York: McGraw Hill.
- Bortz, J. 1977. Lehrbuch der Statistik. Berlin: Springer.
- Bredenkamp, J. 1970. über Maße der praktischen Signifikanz. Zeitschrift für Psychologie, 177, 275-285.
- Bredenkamp, J. 1980. Theorie und Planung psychologischer Experimente. Darmstadt: Steinkopff.
- Burkholder, D. L. 1978. Estimation: Point estimation. In: Kruskal, W. H. & Tanur, J. M. (Hrsg.), 251-259.
- Campbell, D. T. & Stanley, J. C. 1966. Experimental and quasi-experimental designs for research. Chicago: Rand McNally.
- Carmer, S. G. & Swanson, M. R. 1973. Evaluation of ten pairwise multiple comparison procedures by Monte-Carlo methods. Journal of the American Statistical Association, 68, 66-74.
- Cohen, J. 1968. Multiple regression as a general data analytic system. Psychological Bulletin, 70, 426-443.
- Cohen, J. & Cohen, P. 1975. Applied multiple regression/correlation analysis for the behavioral sciences. Hillsdale, N. J.: Lawrence Erlbaum Associates.
- Cramer, E. M. & Nicewander, W. A. 1979. Some symmetric, invariant measures of multivariate association. Psychometrika, 44, 43-54.
- Edwards, A. L. 1971. Versuchsplanung in der psychologischen Forschung. Weinheim: Beltz.
- Edwards, A. L. 1976. An introduction to linear regression and correlation. San Francisco: Freeman.
- Einot, J. & Gabriel, K. R. 1975. A study of the powers of several methods of multiple comparisons. Journal of the American Statistical Association, 70, 574-583.
- Ezekiel, M. & Fox, K. A. 1959. Methods of correlation and regression analysis. New York: Wiley.
- Finn, J. D. 1974. A general model for multivariate analysis. New York: Holt, Rinehart and Winston.
- Fisher, R. A. 1925. Statistical methods for research workers. London: Oliver & Boyd.
- Fisher, R. A. 1935. The design of experiments. London: Oliver & Boyd.
- Fraser, D. A. S. 1979. Inference and linear models. New York: McGraw Hill.
- Gabriel, K. R. 1964. A procedure for testing the homogeneity of all sets of means in analysis of variance. Biometrics, 20, 459-477.
- Gabriel, K. R. 1968. Simultaneous test procedures in multivariate analysis of variance. Biometrika, 55, 489-504.

- Gabriel, K. R. 1969a. Simultaneous test procedures - some theory of multiple comparisons. *Annals of Mathematical Statistic*, 40, 224-250.
- Gabriel, K. R. 1969b. A comparison of some methods of simultaneous inference in MANOVA. In: Krishnaiah, P. R. (Hrsg.), *Multivariate Analysis II*. New York: Academic Press.
- Gaensslen, H. & Schubö, W. 1976. Einfache und komplexe statistische Analyse. München: Reinhardt.
- Garnes, P. A., 1971a. Multiple comparisons of means. *American Educational Research Journal*, 8, 531-65.
- Garnes, P. A., 1971b. Inverse relation between the risks of type I and type II errors and suggestions for the unequal n case in multiple comparisons. *Psychological Bulletin*, 75, 97-102.
- Garnes, P. A., 1973. Type IV errors revisited. *Psychological Bulletin*, 80, 304-307.
- Gocka, E. F. 1973. Regression analysis of proportional cell data. *Psychological Bulletin*, 80, 25-27.
- Graybill, F. A. 1961. An introduction to linear statistical models. Vol. I. New York: McGraw Hill.
- Graybill, F. A. 1976. Theory and application of the linear model. Belmont: Wadsworth.
- Harter, H. L. 1974a. The method of least squares and some alternatives. Part I. *International Statistical Review*, 42, 147-174.
- Harter, H. L. 1974b. The method of least squares and some alternatives. Part II. *International Statistical Review*, 42, 235-264.
- Harter, H. L. 1975a. The method of least squares and some alternatives. Part III. *International Statistical Review*, 43, 1-44.
- Harter, H. L. 1975b. The method of least squares and some alternatives. Part IV. *International Statistical Review*, 43, 125-190.
- Harter, H. L. 1975c. The method of least squares and some alternatives. Part V. *International Statistical Review*, 43, 269-278.
- Hays, W. L. 1973. Statistics for the social sciences. London: Holt, Rinehart and Winston.
- Heck, D. L. 1960. Charts of some upper percentage points of the distribution of the largest characteristic root. *Annals of Mathematical Statistics*, 31, 625-642.
- Herr, D. G. & Gaebelein, J. 1978. Nonorthogonal two-way analysis of variance. *Psychological Bulletin*, 85, 207-216.
- Hild, C. 1977. Schätzen und Testen in einem Regressionsmodell mit stochastischen Koeffizienten. Meisenheim: Hain.
- Horton, R. L. 1978. The general linear model. Data analysis in the social and behavioral sciences. New York: McGraw Hill.
- Hotelling, H. 1951. A generalized T-test and measure of multivariate dispersion. Pro-

- ceedings of the Second Berkeley Symposium on Mathematics and Statistics, 23-41.
- Humak, K. M. S. 1977. Statistische Methoden der Modellbildung. Berlin: Akademie-Verlag.
- Johnson, N. L. & Kotz, S. 1972. Distributions in statistics: Continuous multivariate distributions. New York: Wiley.
- Johnston, J. 1971. Econometric methods. Tokyo: Kogakusha, McGraw Hill.
- Kirk, R. E. 1972. Classification of ANOVA designs. In: Kirk, R. E. 1972. Statistical issues: A reader for the behavioral sciences. Monterey: Wadsworth.
- Kirk, R. E. 1978. Experimental design: Procedures for the behavioral sciences. Belmont, Cal.: Wadsworth.
- Köpcke, W. & Überla, K. (Hrsg.). 1980. Biometrie - heute und morgen. Berlin: Springer.
- Kowalsky, H. J. 1969. Lineare Algebra. Berlin: de Gruyter.
- Krafft, O. 1978. Lineare statistische Modelle. Göttingen: Vandenhoeck und Ruprecht.
- Krishnaiah, R. 1980. Handbook of Statistics, Vol. 1, Analysis of Variance. Amsterdam: North Holland.
- Kruskal, W. H. & Tanur, J. M. (Hrsg.). 1978. International encyclopedia of statistics. New York: The Free Press.
- Lawley, D. N. 1938. A generalization of Fisher's z-test. *Biometrika*, 30, 180-187.
- Leonhard, T. 1975. A Bayesian approach to the linear model with unequal variances. *Technometrics*, 17, 95-102.
- Lindeman, R. H., Merenda, P. F. & Gold, R. Z. 1980. Introduction to bivariate and multivariate analysis. Glenview: Scott, Foresman & Company.
- Lindley, D. V. & Smith, A. F. M. 1972. Bayes estimates for the linear model (with discussion). *Journal of the Royal Statistical Society, Series B*, 34, 1-41.
- Mendenhall, W. 1968. Introduction to linear models and the design and analysis of experiments. Belmont: Wadsworth.
- Mijares, T. A. 1964. Percentage points of the sum  $V_1^{(s)}$  of s roots ( $s = 1-50$ ). Manila: University of the Phillipines, Statistical Center.
- Miller, R. G. 1966. Simultaneous statistical inference. New York: McGraw Hill.
- Moosbrugger, H. 1978. Multivariate statistische Analyseverfahren. Stuttgart: Kohlhammer.
- Moosbrugger, H. 1982. Modelle zur Beschreibung statistischer Zusammenhänge in der psychologischen Forschung. In: Bredenkamp, J. & Feger, H. (Hrsg.), Enzyklopädie der Psychologie, Forschungsmethoden der Psychologie, Bd. 4, Göttingen: Hogrefe.
- Morrison, D. F. 1967. Multivariate statistical methods. New York: McGraw Hill.
- Müller, P. H. (Hrsg.). 1975. Lexikon der Stochastik. Berlin: Akademie-Verlag.

- Namboodiri, N. K., Carter, L. F. & Blalock, H. M. jr. 1975. Applied multivariate analysis and experimental designs. New York: McGraw Hill.
- Olson, C. L. 1976. On choosing a test statistic in multivariate analysis of variance. Psychological Bulletin, 83, 579-586.
- Overall, J. E. & Klett, C. J. 1972. Applied multivariate analysis. New York: McGraw Hill.
- Overall, J. E. & Spiegel, D. K. 1973a. Comment on „Rawling's nonorthogonal analysis of variance“. Psychological Bulletin, 79, 164-167.
- Overall, J. E. & Spiegel, D. K. 1973b. Comment on „Regression analysis of proportionate cell date“. Psychological Bulletin, 80, 28-30.
- Overall, J. E., Spiegel, D. K. & Cohen, J. 1975. Equivalence of orthogonal and nonorthogonal analysis of variance. Psychological Bulletin, 82, 182-186.
- Pearson, E. S. & Wishart, J. 1943. „Student's“ collected papers. London: University College, Biometrika Office.
- Petrinovich, L. F. & Hardyck, C. D. 1969. Error rates for multiple comparison methods. Psychological Bulletin, 71, 43-54.
- Pillai, K. C. S. 1960. Statistical tables for tests of multivariate hypotheses. Manila: University of the Phillipines, Statistical Center.
- Pillai, K. C. S. & Mijares, T. A. 1959. On the moments of the trace of a matrix and approximations to its distribution. Annals of Mathematical Statistics, 30, 1135-1140.
- Pillai, K. C. S. & Jayachandran, K. 1967. Power comparisons of tests of two multivariate hypotheses based on four criteria. Biometrika, 54, 195-210.
- Pillai, K. C. S. & Jayachandran, K. 1970. On the exact distribution of Pillai's  $V^{(s)}$  criterion. Journal of the American Statistical Association, 65, 447-454.
- Rao, C. R. 1973. Linear statistical inference and its applications. New York: Wiley.
- Rao, C. R. & Mitra, S. K. 1971. Generalized inverse of matrices and its application. New York: Wiley.
- Rawlings, R. R. jr. 1972. Note on nonorthogonal analysis of variance. Psychological Bulletin, 77, 373-374.
- Rawlings, R. R. jr. 1973. Comments on the Overall and Spiegel paper. Psychological Bulletin, 79, 168-169.
- Roy, S. N. 1939. p-statistics or some generalisations in analysis of variance appropriate to multivariate Problems. Sankhya, 4, 381-396.
- Roy, S. N. 1946. Multivariate analysis of variance: the sampling distribution of the p-statistics on the non-null hypothesis. Sankhyä, 8, 15-52.
- Roy, S. N. 1957. Some aspects of multivariate analysis. New York: Wiley.
- Ryan, T. A. 1959. Multiple comparisons in psychological Research. Psychological Bulletin, 56, 26-47.
- Schach, S. & Schäfer, T. 1978. Regressions- und Varianzanalyse. Berlin: Springer.

- Scheffé H. 1953. A method for judging all contrasts in the analysis of variance. *Biometrika*, 40, 87-104.
- Scheffé H. 1959. The analysis of variance. New York: Wiley.
- Searle, S. R. 1966. Matrix algebra for the biological sciences. New York: Wiley.
- Searle, S. R. 1971. Linear models. New York: Wiley.
- Seber, G. A. F. 1980. The linear hypothesis: a general theory. London: Griffin and Company.
- Shaffer, J. P. & Gillo, M. W. 1974. A multivariate extension of the correlation ratio. *Educational and Psychological Measurement*, 34, 521-524.
- Späth, H. 1973. Algorithmen für elementare Ausgleichsmodelle. München: Oldenbourg.
- Späth, H. 1974. Algorithmen für multivariable Ausgleichsmodelle. München: Oldenbourg.
- Stanley, J. C. 1978. Linear hypothesis: analysis of variance. In: Kruskal, W. H. & Tanur, J. M. (Hrsg.), 541-554.
- Stevens, J. P. 1972. Global measures of association in multivariate analysis of variance. *Multivariate Behavioral Research*, 7, 37-38.
- Stevens, J. 1979. Comment on Olson: choosing a test statistic in multivariate analysis of variance. *Psychological Bulletin*, 86, 355-360.
- Steyer, R. 1979. Untersuchungen zur nonorthogonalen Varianzanalyse. Weinheim: Beltz.
- Student (Gosset, W. S.). 1908. The probable error of a mean. In: Pearson, E. S. & Wishart, J. 1943. „Student's“ collected papers. London: University College, Biometrika Office.
- Tatsuoka, M. M. 1971. Multivariate analysis: techniques for educational and psychological research. New York: Wiley.
- Timm, N. H. 1975. Multivariate analysis with applications in education and psychology. Monterey: Brooks/Cole.
- Timm, N. H. & Carlson, J. E. 1975. Analysis of variance through full rank models. *Multivariate Behavioral Research*: Monograph 75-1.
- Van de Geer, J. P. 1971. Introduction to multivariate analysis for the social sciences. San Francisco: Freeman.
- Ward, J. H. & Jennings, E. 1973. Introduction to linear models. Englewood Cliffs: Prentice-Hall.
- Wilks, S. S., 1932. Certain generalizations in the analysis of variance. *Biometrika*, 24, 471-494.
- Winer, B. J. 1971. Statistical principles in experimental design. New York: McGraw Hill.
- Wottawa, H. 1981. Probleme der Abtestung der Verteilungsvoraussetzung im allgemeinen linearen Modell. Vortrag auf der 23. Tagung experimentell arbeitender Psychologen, Berlin 1981.
- Zacks, S. 1971. The theory of statistical inference. New York: Wiley.

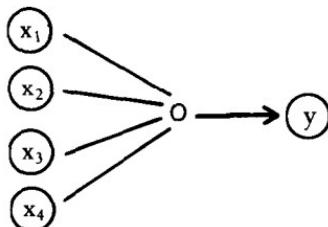
#### 4. Kapitel

## Regressions- und kanonische Analyse

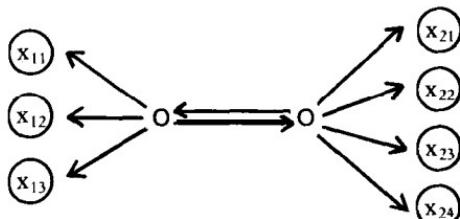
*Werner Schubö, Klaus Haagen und Walter Oberhofer*

Im Rahmen des allgemeinen linearen Modells ist das Ziel der Regressions- und kanonischen Analyse, Zusammenhänge zwischen quantitativen Variablen zu beschreiben und inferenzstatistisch zu deuten. Die Regressionsanalyse beschäftigt sich mit dem Zusammenhang zwischen einer einfachen Variablen und einer Gruppe von Variablen. Die kanonische Analyse beschreibt hingegen den Zusammenhang zwischen zwei Gruppen von Variablen und kann als multiva-

Regressionsanalyse



Kanonische Analyse



Verallgemeinerte kanonische Analyse

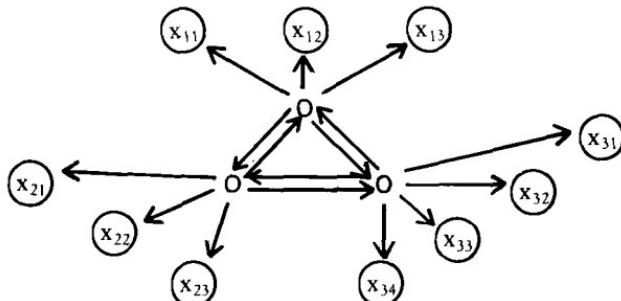


Abb. 1: Die Struktur der Beziehungen bei Regressionsanalyse, kanonischer Analyse und verallgemeinerter kanonischer Analyse.

riate Verallgemeinerung der Regressionsanalyse aufgefaßt werden. Die kanonische Analyse selbst wiederum wird zur kanonischen Analyse für mehrere Variablengruppen verallgemeinert (s. Abb. 1).

Entsprechend der unterschiedlichen Komplexität einzelner Fragestellungen variieren auch die Anforderungen an die Einsatzbereitschaft und die mathematisch-statistischen Vorkenntnisse des Lesers beträchtlich. Da in diesem Beitrag eine systematische Übersicht über das behandelte Gebiet gegeben werden soll, lassen sich schwerer lesbare Passagen nicht vermeiden. Es werden hauptsächlich die Voraussetzungen und Ergebnisse der statistischen Ansätze besprochen; was die mathematischen Ableitungen betrifft, begnügen wir uns mit Literaturhinweisen.

Im Kap. 1 (Regressionsanalyse) wird wohl der Abschnitt 1.8 (Modelle mit Fehlern in den Prädiktoren) am schwierigsten sein. Die Abschnitte 1.10 (Suppression und Kollinearität), 1.11 (Schrittweise Regression) und 1.12 (Teststärke) enthalten nicht nur in der Darstellung, sondern auch im Inhalt bisher nicht beachtete und teils auch nicht bekannte Gesichtspunkte zur Anlage und Interpretation von Regressionsanalysen; sie sind deswegen bewußt besonders leicht lesbar abgefaßt.

Das Kap. 2 (Kanonische Analyse) wird vor allem im Abschnitt 2.6 (Kanonische Analyse für mehrere Variablengruppen) recht komplex. Als Strategie für mathematisch weniger bewanderte Leser sei vorgeschlagen, bei der ersten Lektüre alle nicht unmittelbar verständlichen Formeln zu ignorieren und nur die eingestreuten inhaltlichen Betrachtungen zur Kenntnis zu nehmen. Wie auch in anderen Gebieten der Statistik ist bei diesen Kapiteln nicht zuletzt wegen des mathematischen Aufwandes ein intensives Studium erforderlich, um die hier dargestellten Methoden auch in der Praxis sinnvoll einsetzen zu können.

Zur Notation sei angemerkt, daß Zufallsvariablen und Realisationen von Zufallsvariablen gleich bezeichnet werden, wobei aus der sprachlichen Formulierung des zugehörigen Textes hervorgeht, was gemeint ist. Vektoren werden durch kleine fette Buchstaben, Matrizen durch große fette Buchstaben symbolisiert, während skalare Größen (Merkmale, Zufallsgrößen und Konstanten) im normalen, mageren Druck erscheinen. Tritt eine Matrix oder ein Vektor als Index auf, so unterbleibt der Fettdruck.

## 1. *Regressionsanalyse*

Zunächst werde die Gliederung des Kapitels erläutert. Ausgangspunkt ist die beschreibende lineare Regression, die unmittelbar für den multiplen Fall mit mehreren Prädiktoren formuliert wird. In Abschnitt 1.2 wird ein allgemeines Modell für nichtstochastische Prädiktoren eingeführt, das den Rahmen für die

Kapitel 1.3 bis 1.6 und 1.12 absteckt. In Kap. 1.3 und 1.4 werden beste Schätzungen für unbekannte Größen des Modells und beste Prognosen für zukünftige Werte im Kriterium angegeben. Dies umfaßt sowohl Punkt- als auch Bereichsschätzungen. In Kap. 1.5 stellen wir statistische Tests zur Prüfung von Hypothesen über das Modell vor. Darauf folgt ein Abschnitt, der die Ridge-Regression als eine Möglichkeit beschreibt, im Falle angenäherter Kollinearität der Prädiktoren die Regressionskoeffizienten des Modells zuverlässiger zu schätzen. Während Kap. 1.2 ein Modell für nichtstochastische Prädiktoren enthält, wird in Kap. 1.7 das entsprechende Modell für Prädiktoren formuliert, die selbst Zufallsvariablen sind. Mit fehlerhaft gemessenen Prädiktoren beschäftigt sich Kap. 1.8. Als eine Anwendung des allgemeinen Ansatzes aus Kap. 1.2 enthält Kap. 1.9 Hinweise auf die regressionsanalytische Behandlung von Zeitreihen. Hauptsächlich praktische Probleme bei der Planung und Interpretation von regressionsanalytischen Studien stehen bei den drei letzten Abschnitten im Vordergrund.

Die hier nicht oder nicht ausführlich behandelten Themen sollen wenigstens durch Literaturhinweise ersetzt werden. Dazu gehören:

Interpretation einer Computerprogramm-Ausgabe (Nie et al., 1975, p. 358; Gaensslen & Schubö, 1976, p. 319; Schuchart-Ficher et al., 1980, p. 49),

Dummy-Variablen als Prädiktoren zur regressionsanalytischen Berechnung einer Varianzanalyse (Cohen & Cohen, 1975, p. 291-424; Gaensslen & Schubö, 1976, p. 147; Sievers, 1977),

Behandlung von Repeated Measurement Designs (Cohen & Cohen, 1975, p. 413; Edwards, 1979, p. 117),

Regression mit qualitativen und quantitativen Variablen zur Skalierung (Young et al., 1976),

Regression auf Polynome und trigonometrische Polynome (Cohen & Cohen, 1975, p. 213; Graybill, 1976, p. 302),

Nichtlineare Regression (Marquardt, 1963; Bard, 1974; Milliken & Graybill, 1970),

Monotone Regression (Carroll & Chang, 1964; Kruskal, 1971; de Leeuw, 1977; Sulz, 1980, p. 124),

Missing Data-Behandlung (Timm, 1970; Lösel & Wüstendorfer, 1974; Cohen & Cohen, 1975; Frane, 1976),

Inverse Regression und Kalibrierung (Krutchkoff, 1967; Hoadley, 1970),

Stückweise Regression (McGee & Carleton, 1970; Wainer, 1971),

Konfidenzbereiche für Regressionskoeffizienten von nicht-ellipsoidischer Gestalt (Bowden, 1970),

Eingeschränkter Wertebereich als Nebenbedingung an die Regressionskoeffizienten (Lovell & Prescott, 1970),

Regressionsmodelle für die Sterbetafel-Analyse (Cox, 1972).

Kontingenztafelanalyse im linearen Modell nach Grizzle, Starner und Koch und nach Goodman (Küchler, 1979, p. 154-255),

Kovarianzstrukturanalyse und partial least squares (Jöreskog, 1973, 1977, 1978; McDonald, 1974; Lohmöller, 1979),

Pfadanalyse (Seibel & Nygreen, 1972; Vetter, 1972; Hummell & Ziegler, 1976) und Anwendung der Regressionsanalyse zur Lösung des Interpretationsproblems in der mehrdimensionalen Skalierung und Faktorenanalyse (Gigerenzer, 1981, p. 338).

## 1.1 Beschreibende lineare Regression

Zunächst soll nur rein beschreibend die lineare Regression in einer Gesamtheit von Beobachtungen dargestellt werden. Es ist dabei nicht beabsichtigt, die Regressionsbeziehung über die vorliegende Gesamtheit hinaus einer Grundgesamtheit oder einem Modell zuzuordnen. Die hier dargestellte deskriptive Regression versteht sich nur als eine lineare Approximation einer Gesamtheit von Beobachtungen.

Wir gehen davon aus, daß von einer Variablen  $y$  und von weiteren  $K$  Variablen  $x_1, x_2, x_3, \dots, x_K$  jeweils  $N$  Messungen vorliegen. Die Variable  $y$  nennen wir das *Kriterium* oder auch abhängige *Variable*, *Regressand* oder *endogene Variable*. Die Variablen  $x_1, x_2, x_3, \dots, x_K$  werden als *Prädiktoren*, *unabhängige Variablen*, *Regressoren* oder *exogene Variablen* bezeichnet.

In einem Beispiel aus der Wirtschaftspsychologie könnte etwa als Kriterium  $y$  die Produktivität oder auch die Dauer des Arbeitsverhältnisses betrachtet werden, deren Zusammenhang mit den Prädiktoren  $x_k$ : Arbeitsbedingungen, Zufriedenheit mit dem Vorgesetzten, Bezahlung und Vielfältigkeit der Tätigkeitsbereiche untersucht werden könnte. In einem Therapieexperiment zur Reduktion des Rauchens könnten als Kriterium  $y$  der tägliche Zigarettenkonsum und als Prädiktoren  $x_k$  die Variablen Arbeits-/Freizeittag, Ausmaß des eingesetzten Treatments und schließlich die fortlaufende Nummer des Tages dienen. Im letztgenannten Beispiel sind die Ergebniseinheiten oder Merkmalsträger die verschiedenen Tage, an denen eine Person beobachtet wurde, während im ersten Beispiel die Variablen jeweils an verschiedenen Beschäftigten in Form einer Querschnittsuntersuchung erhoben werden.

Bei der linearen Regression ist man bestrebt, den Zusammenhang zwischen dem Kriterium  $y$  und den Prädiktoren  $x_1, x_2, \dots, x_K$  durch

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_K x_K$$

zu beschreiben. Wenn diese Beziehung zutreffen würde, könnte man mit ihrer Hilfe bei gegebenen Werten der Prädiktoren und bei Kenntnis der für die Gesamtheit aller Beobachtungen konstanten Koeffizienten  $\beta_0, \beta_1, \dots, \beta_K$  ein-

deutig die Werte des Kriteriums reproduzieren. In der Empirie ist es jedoch im allgemeinen nicht möglich, die Werte des Kriteriums durch eine Funktion so einfacher Bauart zu reproduzieren. Als Ausweg bieten sich daher zwei Möglichkeiten an. Entweder man sucht eine kompliziertere Funktion, die eine perfekte Reproduktion ermöglicht, oder man betrachtet die sich aus  $\beta_0 + \beta_1 x_1 + \dots + \beta_K x_K$  ergebenden Werte als Approximation für das Kriterium  $y$ . Wir folgen hier dem zweiten Weg und schreiben für diese Approximation von  $y$

$$(1) \quad \hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_K x_K.$$

Verwenden wir also (1) als Approximation für das Kriterium, so ergibt sich als *Approximationsfehler (Residuum)*

$$u = y - \hat{y}.$$

Damit läßt sich der Zusammenhang zwischen dem Kriterium  $y$ , der Approximation  $\hat{y}$  und dem Approximationsfehler  $u$  auch schreiben als

$$(2) \quad y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_K x_K + u.$$

Die Koeffizienten  $\beta_0, \beta_1, \dots, \beta$  sollen nun so bestimmt werden, daß die Approximation von  $y$  durch  $\hat{y}$  möglichst gut ist, oder, was dasselbe besagt, daß der Approximationsfehler  $u$  möglichst klein ist. Diese Aufgabe, deren Lösung im Aufsuchen des Satzes von Koeffizienten besteht, wird in der numerischen Mathematik als Ausgleichung bezeichnet. Eine Voraussetzung zur Lösung ist, daß genauer angegeben wird, was mit „möglichst kleinen Fehlern“ gemeint sein soll.

Man könnte etwa fordern, daß der größte Fehler  $u$  in der betrachteten Gesamtheit dem Betrage nach so klein wie möglich sein soll. Dies führt zum Problem der Ausgleichung nach Tschebyscheff, die z.B. in Stiefel (1961, p. 47) auf einfache Weise beschrieben ist. Die Lösung kann durch die Simplex-Methode, einen Algorithmus aus der linearen Programmierung, erfolgen. Dieser Ansatz hat den Nachteil, daß spätestens dann erhebliche Schwierigkeiten auftauchen, wenn inferenzstatistisch von Stichproben auf Grundgesamtheiten geschlossen werden soll.

Die bislang am weitesten verbreitete und in ihren Folgen am besten untersuchte Forderung ist, die Summe der quadrierten Fehler als Maß der Güte einer Approximation heranzuziehen. Diese Forderung wird *Prinzip der kleinsten Quadrate* oder *Minimum-Quadrat-Prinzip (MQ-Prinzip, LS-Prinzip)* genannt und soll auch hier im Vordergrund stehen. Einerseits ist sie rechentechnisch leichter zu handhaben, und andererseits ist sie für inferenzstatistische Schlüsse besser geeignet.

Zur Formulierung des Minimum-Quadrat-Prinzips soll zunächst eine übersichtliche Schreibweise in Form von Vektoren und Matrizen eingeführt werden. Beispielsweise schreiben wir für die Werte der Variablen  $y$  bzw.  $u$  bzw.  $x_k$  in einer (stets als endlich angenommenen) Gesamtheit vom Umfang  $N$

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} \text{ bzw. } \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{pmatrix} \text{ bzw. } \mathbf{x}_k = \begin{pmatrix} x_{1k} \\ x_{2k} \\ \vdots \\ x_{Nk} \end{pmatrix}.$$

Der Satz von Koeffizienten  $\beta_0, \beta_1, \dots, \beta_K$  bildet den Spaltenvektor

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{pmatrix}.$$

Fassen wir die Vektoren  $x_k$  als Spalten einer Matrix auf, so erhalten wir die Matrix  $X$  der Prädiktorwerte

$$\mathbf{X} = (\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_K) = \begin{pmatrix} x_{10} & x_{11} & \dots & x_{1K} \\ x_{20} & x_{21} & \dots & x_{2K} \\ \vdots & \vdots & & \vdots \\ x_{N0} & x_{N1} & \dots & x_{NK} \end{pmatrix},$$

wobei die Variable  $x_0$  als Multiplikator zu  $\beta_0$  zugunsten einer einheitlichen Schreibweise eingeführt wurde und stets den Wert 1 hat.

Mit dieser Schreibweise wird aus dem linearen Ansatz (2)

$$(3) \quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}.$$

Aus (1) wird

$$(4) \quad \hat{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta}.$$

Die Werte der Restvariablen  $u$  ergeben sich dementsprechend als

$$(5) \quad \mathbf{u} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}.$$

Das Prinzip der kleinsten Quadrate lässt sich nun als

$$(6) \quad S(\boldsymbol{\beta}) = \mathbf{u}'\mathbf{u} = \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \stackrel{!}{=} \text{Minimum}$$

formulieren.

$S(\boldsymbol{\beta})$  ist die Summe der quadrierten Werte der Restvariablen. Die Summe hängt von  $\boldsymbol{\beta}$  ab, was durch die Angabe von  $\boldsymbol{\beta}$  in Klammern hervorgehoben wird.

Natürlich hängt die Summe auch von den Werten  $y_n$  für  $n = 1, 2, \dots, N$  und  $x_{nk}$  für  $k = 1, 2, \dots, K$  und  $n = 1, 2, \dots, N$  in der Gesamtheit ab; diese Werte betrachten wir für die Minimierungsaufgabe jedoch als fest, weshalb ihr Einfluß auf die Summe  $S$  nicht dokumentiert werden soll.

Es läßt sich zeigen, daß notwendige und hinreichende Bedingungen für die Werte des Vektors  $\beta$  mit minimaler Fehlerquadratsumme  $S(\beta)$  die sog. Normalgleichungen

$$(7) \quad X'X\beta = X'y$$

sind. Daß diese Normalgleichungen für (6) notwendige Bedingungen sind, wird in beinahe allen einschlägigen Lehrbüchern (z.B. Johnston, 1963, p. 11, 109; Draper & Smith, 1966, p. 10, 59; Gaensslen & Schubö, 1976, p. 299; Haagen & Pertler, 1976, p. 70; Hummell & Ziegler, 1976, p. 40, 46; Moosbrugger, 1978, p. 51, 59; Edwards, 1979, p. 13, 40; Küchler, 1979, p. 77) gezeigt. Eine vollständige Darstellung erfordert aber auch den Nachweis, daß jede Lösung von (7) auch die Minimum-Quadrat-Forderung (6) erfüllt (siehe z.B. Schönfeld, 1969, p. 30; Wonnacott & Wonnacott, 1970, p. 13; Theil, 1971, p. 34; Rao, 1973, p. 180; Schneeweiß, 1974, p. 43; Graybill, 1976, p. 174, 343; Schach & Schäfer, 1978, p. 12).

Für die Normalgleichungen (7) gibt es stets wenigstens einen Lösungsvektor, dessen Komponenten dann Regressionskoeffizienten heißen. Wenn die Matrix  $X'X$  nichtsingulär (invertierbar, von vollem Rang) ist, gibt es genau eine Lösung

$$(8) \quad \beta = X^+y$$

mit

$$X^+ = (X'X)^{-1}X'.$$

Dies ist für empirische Daten gewöhnlich der Fall.

Es gibt jedoch zwei Situationen, in denen die Matrix  $X'X$  im „wissenschaftlichen Alltag“ singulär werden kann:

1. Die Anzahl der Beobachtungen  $N$  ist zu klein; sie ist nicht größer als die Anzahl der Prädiktoren  $K$ .
2. Es besteht ein exakter linearer Zusammenhang eines Prädiktors mit anderen Prädiktoren. Wenn etwa eine gemessene Variable  $x_k$  konstant ist, besteht ein linearer Zusammenhang zu  $x_0$ . Auch wenn  $x_k$  der summativ Gesamtwert aus den Prädiktoren  $x_1, x_2$  und  $x_3$  ist, besteht eine derartige lineare Abhängigkeit.

Wenn die Matrix  $X'X$  singulär ist, gibt es unendlich viele Lösungen der Normalgleichungen, die alle zur gleichen minimalen Quadratsumme  $S(\beta)$  führen.

Dies kann sich so auswirken, daß ein oder mehrere Regressionskoeffizienten (Anzahl entspricht dem Rangdefekt) frei wählbar sind; in der Regel kann jedoch die Vielfalt der Lösungen nicht so einfach angegeben werden. Wird die Regressionsanalyse durch ein Computerprogramm ausgeführt, äußert sich die Singularität der Matrix durch einen Abbruch der Berechnungen mit der Fehlermeldung „Division durch Null“ oder „Determinante ist Null“, da in den Programmen keine Vorsorge für eine derartige Datenkonstellation getroffen ist. Im allgemeinen wird man dann einige Prädiktoren ausschließen, um wenigstens für eine abgewandelte Regressionsaufgabe eine Lösung zu erhalten. Später werden wir unter den Stichwörtern Kollinearität und Ridge-Regression noch auf ähnliche Situationen zurückkommen, wo  $X'X$  nur „beinahe“ singulär ist; ansonsten soll in Zukunft der in diesem Absatz besprochene Fall ausgeschlossen sein.

Verwendet man im linearen Beschreibungsmodell (3) die Lösung (8) für  $\hat{u}$  aufgrund der Minimum-Quadrat-Forderung (6), lassen sich eine Reihe von Eigenschaften der Restvariablen  $u$  ableiten, ohne daß zusätzliche Annahmen notwendig sind.

So sind die Produktsummen der Residuen mit jeder Prädiktorvariablen Null

$$(9) \quad X'u = 0,$$

woraus sich spezieller ergibt, daß der Mittelwert der Fehlervariablen Null ist, und daß die empirischen Kovarianzen zwischen der Fehlervariablen und allen Prädiktorvariablen verschwinden:

$$(10) \quad \bar{u} = 0$$

$$(11) \quad s_{ux_k} = 0 \text{ für alle } k = 1, \dots, K.$$

Eine weitere Folge von (9) ist

$$(12) \quad \hat{y}'u = 0,$$

weshalb die Kovarianz zwischen der Approximation  $\hat{y}$  und der Fehlervariablen Null ist. Weiterhin gilt die wichtige Zerlegung

$$(13) \quad y'y = \hat{y}'\hat{y} + u'u$$

die auch in Form empirischer Varianzen geschrieben werden kann

$$(14) \quad s_y^2 = s_{\hat{y}}^2 + s_u^2.$$

Hieraus läßt sich die Interpretation des Bestimmtheitsmaßes

$$(15) \quad R^2 = \frac{s_{\hat{y}}^2}{s_y^2} = 1 - \frac{s_u^2}{s_y^2}$$

als Maß der Anpassungsgüte ablesen; es liegt um so näher bei 1, je kleiner die Fehlerwerte sind, während es sich bei großem Fehleranteil der Zahl Null nähert.

Obwohl im Rahmen der beschreibenden Regression noch eine Reihe wichtiger Themen wie polynomische Regression, Bedeutung einzelner Prädiktoren usw. besprochen werden könnte, sollen zunächst die Populationsmodelle für inferenzstatistische Aufgaben dargestellt werden, um den Zusammenhang der Argumentation nicht durch zu viele Einzelheiten zu verwischen.

## 1.2 Das allgemeine regressionsanalytische Modell

Bei den meisten Anwendungen der Regression möchte man zu Aussagen kommen, die über die konkrete Stichprobe hinausreichen. In den anfangs erwähnten Beispielen interessiert eher der Zusammenhang zwischen der Produktivität  $y$  und den Prädiktoren Arbeitsbedingungen, Zufriedenheit usw. bei *zukünftig* Beschäftigten bzw. das Ausmaß des Rauchens bei *zukünftiger* Anwendung von Treatments als die Zusammenhänge in der Stichprobe. Dies entspricht der Situation von Galileo Galilei, der 1604 sein Fallgesetz  $v = gs$  mit einem universellen Anspruch auch für zukünftig fallende Massen aufstellte. Die Form des Fallgesetzes sowie die nähere Kennzeichnung der Umstände, unter denen es zutreffen soll, bilden ein Populationsmodell (= Satz von Oberhypothesen) mit dem freien Parameter Fallbeschleunigung  $g$ , der dann durch einige Beobachtungen näher eingegrenzt werden kann. Das Populationsmodell sollte natürlich auch durch Beobachtungen begründet und überprüft werden. Obwohl wir vornehmlich Aussagen über Parameter des Modells treffen wollen, werden wir auch einige Möglichkeiten zur Überprüfung des Modells kennenlernen. übrigens erwies sich auch Galileis ursprüngliches Modell als revisionsbedürftig, da es dem später allgemein verwendeten Fallgesetz  $v = gt$  unterlegen war. Außer dem in diesem Kapitel einzuführenden regressionsanalytischen Modell wird in Kap. 1.7 noch das korrelationsanalytische Modell vorgestellt.

Für die Formulierung der Populationsmodelle wird auf den Begriff der Zufallsvariablen mit zugehörigen Verteilungsfunktionen (2. B. Haagen & Pertler, 1976, p. 169) zurückgegriffen. Wir wollen nämlich annehmen, daß die Beobachtungen in Stichproben als Realisierungen von Zufallsvariablen aufgefaßt werden können.

Das allgemeine *regressionsanalytische Modell* (*general linear model, GLS-Modell*) umfaßt folgende Annahmen:

(A1) Es besteht die Beziehung

$$y = X\beta + u,$$

wobei  $X$  eine  $N \times (K+1)$ -Matrix von festen, beobachtbaren

Werten,  $\beta$  ein Vektor von festen, unbekannten Parametern mit  $K+1$  Komponenten,

$y$  ein Vektor von  $N$  beobachtbaren Zufallsgrößen und  $u$  ein Vektor von  $N$  unbeobachtbaren Zufallsgrößen sind.

(A2) Die Matrix  $X$  ist vom Rang  $K+1$

(A3) Der Erwartungswert von  $u$  ist der Nullvektor:  $Eu = 0$

(A4) Die Kovarianzmatrix  $\Sigma_{uu}$  zu  $u$  existiert und ist nichtsingulär.

$\Sigma_{uu}$  läßt sich als  $\Sigma_{uu} = \sigma^2 V$  mit  $\sigma^2 = \frac{1}{N} \text{Spur } \Sigma_{uu}$  darstellen.

(A5) Der Vektor  $u$  ist  $N$ -dimensional normalverteilt.

Diese Annahmen bedürfen der Erläuterung.

In A1 wird die Art der in der Regression beteiligten Größen und ihr Zusammenhang eingeführt. Jede der  $N$  Zeilen von  $X$  umfaßt die  $K+1$  Werte der Prädiktorvariablen eines Merkmalsträgers. Ein Merkmalsträger zur Zeile  $n$  soll eine Ziehung aus der Grundgesamtheit der Merkmalsträger zu den zugehörigen festen Werten  $x_{n0}, x_{n1}, \dots, x_{nK}$  der Prädiktorvariablen sein. Mit der Grundgesamtheit von Merkmalsträgern ist zugleich eine Wahrscheinlichkeitsverteilung der  $n$ -ten Komponente  $y_n$  der Kriteriumsvariablen durch die  $n$ -te Komponente  $u$ , der Fehlervariablen verknüpft. Die Realisierungen von  $y_n$  und  $u_n$  und damit auch ihre Wahrscheinlichkeitsverteilungen unterscheiden sich nur additiv durch den festen Wert der  $n$ -ten Komponente  $(X\beta)_n$ .

Eine wichtige Konsequenz der Annahme, daß die Werte von  $X$  fest sind, ist, daß sich die Interpretation der Wahrscheinlichkeiten als relative Häufigkeiten für sehr große Anzahlen von Versuchswiederholungen auf wiederholte Ziehungen zu denselben Prädiktorwerten stützt. Das bedeutet allerdings nicht, daß sich Inferenzen nur auf die betrachteten Prädiktorwerte beziehen können; Inferenzen sind stets in dem Bereich von Prädiktorwerten möglich, in dem die Modellannahmen gültig sind (Graybill, 1976, p. 147). Die Annahme fester Werte in  $X$  verlangt vor allem, daß die Prädiktoren fehlerfrei gemessen werden. Bei kontinuierlichen Prädiktorvariablen ist diese Forderung natürlich nie erfüllbar, weshalb man sich dort mit der Annahme kleiner relativer Meßfehler zufrieden geben muß; inferenzstatistische Schlüsse bleiben dann nur noch näherungsweise gültig.

Durch die Annahme A2 wird lediglich der unbequeme Fall der Nicht-Identifizierbarkeit des Vektors  $\beta$  ausgeschlossen. Wäre  $X$  nicht vom Rang  $K+1$ , gäbe es beliebig viele verschiedene Vektoren  $\beta$ , die unter sonst gleichen Umständen zu derselben Verteilung des Zufallsvektors  $y$  führen. A2 ist der Forderung gleichwertig, daß  $X'X$  nichtsingulär ist. Die Forderung kann immer durch Umorganisation der Prädiktorvariablen und Verkleinerung ihrer Anzahl erfüllt werden, ohne daß der Erklärungswert der Prädiktoren geschmälert würde.

Die Annahme A3 besagt, daß für jede betrachtete Konstellation der Prädiktorwerte der Erwartungswert („long-run-Mittelwert“) der Fehlervariablen Null ist, was damit gleichwertig ist, daß der Erwartungswert der Kriteriumsvariablen

$$(16) \quad E\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$$

linear durch die Prädiktoren bestimmt ist. Es besteht eine Ähnlichkeit von A3 mit der Eigenschaft (10) der Minimum-Quadrat-Methode, daß der Mittelwert des Approximationsfehlers verschwindet. Eine spezielle Minimum-Quadrat-Forderung benötigt man wegen Annahme A3 nicht.

In Annahme A4 wird zunächst lediglich festgelegt, daß die Varianzen der Fehlerwerte und ihre Kovarianzen zu verschiedenen Prädiktorkombinationen existieren. Damit werden z.B. Autokorrelationen der Fehlerwerte  $\text{cov}(u_n, u_m) \neq 0$  für  $n \neq m$  oder Heterogenität der Fehlervarianzen  $\text{var } u_n \neq \text{var } u_m$  für  $n \neq m$  zugelassen. Weil manchmal zwar die absoluten Größen der Varianzen und Kovarianzen unbekannt, aber ihre Verhältnisse untereinander bekannt sind, wird meist die Kovarianzmatrix in die Faktoren  $\sigma^2$  und V zerlegt. Wir behandeln hier nur den Fall, daß V nichtsingulär ist (sonst s. Theil, 1971, p. 274). Man fordert dann, daß die NxN-Matrix V bekannt sein muß, während  $\sigma^2$  als einzelner unbekannter Parameter geschätzt wird. Beispielsweise wird im *klassischen regressionsanalytischen Modell (ordinary least squares model, OLS-Modell)*, das die meisten Computerprogramme zur Regression voraussetzen, als Matrix V die Einheitsmatrix I vorausgesetzt:

$$(K4) \quad \Sigma_{uu} = \sigma^2 \mathbf{I},$$

was Unkorreliertheit der Fehlervariablen und  
Varianzhomogenität (Homoskedastizität) bedeutet.

Das klassische regressionsanalytische Modell ist also ein Spezialfall des allgemeinen regressionsanalytischen Modells.

Die Annahme A5 schließlich fügt eine Voraussetzung über die Form der Verteilung der Kriteriumsvariablen für feste Prädiktorwerte hinzu. Diese Annahme ist für die Beurteilung der Eigenschaften der noch zu besprechenden Schätzfunktionen nicht erforderlich, sondern wird erst bei der Konstruktion von Konfidenzbereichen und statistischen Tests benötigt. Allerdings wird für Schätzer nach dem *Maximum-Likelihood-Prinzip (ML-Prinzip)* (z.B. Schönfeld, 1969, p. 110; Graybill, 1976, p. 75) bereits bei der Konstruktion des Schätzers der Typ der Wahrscheinlichkeitsverteilung von u vorausgesetzt, wobei wir annehmen, daß die Wahrscheinlichkeitsverteilung parametrisierbar ist.

### 1.3 Die Schätzung der Parameter im allgemeinen regressionsanalytischen Modell

Wenn man weiß, daß ein allgemeines regressionsanalytisches Modell mit den Annahmen A1 bis A4 vorliegt, und wenn eine Stichprobe vom Umfang N mit den Prädiktorwerten X und den Kriteriumswerten y vorliegt, dann besteht die Aufgabe darin, die noch unbekannten Parameter  $\beta$  und  $\sigma^2$  einzugrenzen. Die Matrix V aus A4 sei bekannt. Zunächst sollen Punktschätzungen und ihre Eigenschaften angegeben werden.

Der *beste (effiziente, kleinste Varianz) lineare erwartungstreue* (kein Bias) Schätzer für  $\beta$ , den man ohne a priori-Information über  $\beta$  oder  $\sigma^2$  gewinnen kann (Seifert, 1975, p. 44), ist

$$(17) \quad \hat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y},$$

der sog. *GLS-Schätzer (generalized least squares estimator)*. Wenn die Voraussetzung der Normalverteilung A5 hinzugenommen wird, ist dieser Schätzer sogar in der Klasse der erwartungstreuen Schätzer (linear oder nichtlinear) effizient (Theil, 1971, p. 390). Unter dieser Voraussetzung ist  $\beta$  zugleich auch der Maximum-Likelihood-Schätzer.

Wendet man das Minimum-Quadrat-Prinzip auf die Daten der Stichprobe an, dann erhält man entsprechend Gleichung (8) den sogenannten *OLS-Schätzer (ordinary least squares estimator)*:

$$(18) \quad \mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y},$$

der unter der zusätzlichen Annahme K4 des klassischen regressionsanalytischen Modells dem GLS-Schätzer gleich ist. Auch der OLS-Schätzer ist ein linearer erwartungstreuer Schätzer (Johnston, 1963, p. 188), der aber nur im *klassischen* regressionsanalytischen Modell kleinste Varianz hat. Der Effizienzverlust im *allgemeinen* regressionsanalytischen Modell ist für Punktschätzungen jedoch nicht groß (Wonnacott & Wonnacott, 1970, p. 334), so daß die Verwendung des OLS-Schätzers vertretbar scheint.

übrigens kann auch der GLS-Schätzer  $\hat{\beta}$  im allgemeinen regressionsanalytischen Modell nach einem abgewandelten Minimum-Quadrat-Prinzip konstruiert werden. Man erhält ihn, indem man das Minimum von

$$(\mathbf{y}-\hat{\mathbf{y}})' \mathbf{V}^{-1} (\mathbf{y}-\hat{\mathbf{y}}) = \sum_{k=1}^K \sum_{l=1}^K (y_k - \hat{y}_k) (y_l - \hat{y}_l) (\mathbf{V}^{-1})_{kl}$$

$$\text{mit } \hat{\mathbf{y}} = \mathbf{X}\hat{\beta}$$

für  $\beta$  sucht. Deshalb wird der GLS-Schätzer häufig auch als WLS-Schätzer (weighted least squares) bezeichnet.  $\hat{y}$  ist die Schätzung des Erwartungswertes  $Ey = X\beta$ .

Ein erwartungstreuer quadratischer (z.B. Schönfeld, 1969, p. 67) Schätzer für  $\sigma^2$  ist

$$(19) \quad \hat{\sigma}^2 = \frac{1}{N-K-1} (y - \hat{y})' V^{-1} (y - \hat{y})$$

oder gleichwertig

$$\hat{\sigma}^2 = \frac{1}{N-K-1} y' (V^{-1} - V^{-1} X (X' V^{-1} X)^{-1} X' V^{-1}) y.$$

Dieser Schätzer ist unter der zusätzlichen Annahme der Normalverteilung A5 effizient in der Klasse der erwartungstreuen quadratischen Schätzer. Der ML-Schätzer für  $\sigma^2$  weicht von (19) etwas ab, ist allerdings auch nicht erwartungstreu. Im klassischen regressionsanalytischen Modell entfällt  $V^{-1}$  (= Einheitsmatrix), und  $\hat{\sigma}^2$  ist die Schätzung der Störvarianz.

Durch eine *Bereichsschätzung* wird ein Bereich um die Punktschätzung angegeben, der mit der Wahrscheinlichkeit  $1-a$  den zu schätzenden Parameter enthält (Haagen & Seifert, 1979, p. 140). Die Wahrscheinlichkeit  $a$ , daß der Bereich den Parameter nicht überdeckt, wird Irrtumswahrscheinlichkeit genannt. Der Konfidenzbereich wird aufgrund der Werte des Zufallsvektors  $y$  festgelegt und ist insofern von der Realisierung der Zufallsvariablen abhängig. Aus der Bereichsschätzung kann auch ein statistischer Test zur Prüfung einer Hypothese über den Parameter abgeleitet werden.

Zunächst sollen Bereichsschätzungen für die Regressionskoeffizienten  $\beta$  eingeführt werden. Um das Spektrum der Anwendungen zu vergrößern, wird die Bereichsschätzung allgemeiner für beliebige Linearkombinationen

$$(20) \quad \gamma = L\beta$$

mit fester  $M \times (K+1)$ -Matrix  $L$   
mit  $M \leqq K+1$  und Rang  $L = M$

durchgeführt. Die Matrix  $L$  wird auch als *Kontrast* bezeichnet. Setzt man die Einheitsmatrix für  $L$  ein, ist  $\gamma = \beta$ . Man weiß übrigens (Seifert, 1975, p. 57), daß die beste lineare erwartungstreue Schätzung von  $\gamma$  durch

$$(21) \quad \hat{\gamma} = L\hat{\beta}$$

aus dem GLS-Schätzer für  $\beta$  berechnet werden kann.

Als Konfidenzbereich für  $\gamma$  zum Konfidenzgrad  $1-\alpha$  wird der Bereich für  $\gamma$  als

$$(22) \quad \frac{1}{M\hat{\sigma}^2} (\gamma - \hat{\gamma})' (L(X'V^{-1}X)^{-1}L')^{-1} (\gamma - \hat{\gamma}) \leq F_{\alpha}(M, N-K-1)$$

konstruiert. Der angegebene Bereich ist im allgemeinen das Innere eines  $M$ -dimensionalen Ellipsoids mit Mittelpunkt  $\hat{\gamma}$ , für den Fall nur einer einkomponentigen Linearkombination ergibt sich ein Intervall. Ein wichtiger Spezialfall ist das Konfidenzintervall für einen einzelnen Regressionskoeffizienten  $\beta_k$  im klassischen regressionsanalytischen Modell ( $V = 1$ ), das man durch den Ansatz  $L = (0,0, \dots, 0,1,0, \dots, 0)$  erhält:

$$(23) \quad \frac{(\beta_k - b_k)^2}{\hat{\sigma}_{b_k}^2} \leq F_{\alpha}(1, N-K-1)$$

$$\text{mit } \hat{\sigma}_{b_k}^2 = \sigma^2 [(X'X)^{-1}]_{kk};$$

die Größe  $\sqrt{\hat{\sigma}_{b_k}^2}$  wird bei Computerprogrammen zur Regression unter dem Titel Standardfehler des Regressionskoeffizienten ausgedruckt. Eine Bereichsschätzung für den Parameter  $\sigma^2$ , der im klassischen Regressionsmodell die Varianz der Fehlervariablen  $u_k$  ist, kann bei einem Signifikanzniveau von  $\alpha$  durch

$$\frac{(N-K-1) \hat{\sigma}^2}{\chi_{1-\alpha/2}^2(N-K-1)} \leq \sigma^2 \leq \frac{(N-K-1) \hat{\sigma}^2}{\chi_{\alpha/2}^2(N-K-1)}$$

angegeben werden (Schneeweiß, 1974, p. 65, 111).

## 1.4 Prognose im allgemeinen regressionsanalytischen Modell

Im vorangehenden Kapitel hatten wir diskutiert, wie in einem bestimmten Modell (A1 bis A5, K4) aus den Werten einer Stichprobe die unbekannten Parameter der Verteilung geschätzt werden können. Durch die Schätzung erhalten wir empirische Aussagen über den Zusammenhang von Prädiktor- und Kriterienwerten. Wenn Art und Ausmaß des Zusammenhangs für zukünftige Beobachtungen weiter gelten, können wir versuchen, zukünftige Werte des Kriteriums aus denen der Prädiktoren vorherzusagen. In dem am Anfang von Kap. 1.1 geschilderten Beispiel wollen wir also wissen, welchen Wert der Produktivität wir für zukünftige Beschäftigte vermuten, wenn wir ihre Arbeitsbedingungen, die Zufriedenheit mit dem Vorgesetzten, die Bezahlung und die Vielfältigkeit der Tätigkeitsbereiche kennen. Die so beschriebene Fragestellung der Prognose ist im Grunde ebenfalls eine Aufgabe der Schätzung, wobei im Gegensatz zum letzten Kapitel nicht feste Parameter einer Verteilung, sondern die Werte einer Zufallsvariablen geschätzt werden sollen.

Nach dem Ziel der Prognose können passive und aktive Prognose unterschieden werden. Von der passiven Prognose soll bei einer festen, von außen vorgegebenen zukünftigen Konstellation der Prädiktoren abgeschätzt werden, wie die Werte des Kriteriums liegen werden. In der aktiven Prognose werden verschiedene alternative zukünftige Konstellationen der Prädiktoren in ihrer Auswirkung auf das Kriterium untersucht; das betrachtete System wird dann der Prädiktorkonstellation ausgesetzt, die den nach externen Beurteilungsmaßstäben günstigsten Wert des Kriteriums erwarten läßt. Die aktive Prognose ist etwa auf Verlaufsuntersuchungen von klinisch-psychologischen Therapien anwendbar, wo Treatments im Hinblick auf günstigen Symptomverlauf ausgewählt werden sollen. Auch bei der Auswahl von Bewerbern durch Regression wird aktive Prognose verwendet. Nur wenn die Prädiktorkonstellation nicht beeinflußt werden kann, wenn sie etwa ihrerseits prognostiziert werden muß, kann nur passiv prognostiziert werden. Die darzustellenden Prognoseansätze sind sowohl für aktive wie passive Prognose verwendbar.

Zusätzlich zu den Beobachtungen der Zufallsvariablen  $y_n$  für  $n = 1, 2, \dots, N$  und der festen Variablenwerte  $x_{nk}$  für  $n = 1, 2, \dots, N$  und  $k = 0, 1, \dots, K$  seien  $M$  neue Merkmalsträger mit den Werten für die Prädiktorvariablen  $x_{nk}$  für  $n = N+1, N+2, \dots, N+M$  und  $k = 0, 1, \dots, K$  beobachtet, ohne daß zugleich die Werte der Kriteriumsvariablen bekannt wären. Die Prädiktorwerte fassen wir in der  $M \times (K+1)$ -Matrix  $X_z$  und die zugehörigen Zufallsvariablen  $y_n$  im Vektor  $y_z$  mit  $M$  Komponenten zusammen. Die zusätzlichen Prädiktorwerte seien fest, d.h., daß sie fehlerfrei gemessen und nicht zufällig sind. Für viele Anwendungen genügt es, die Prognose nur für einen weiteren Merkmalsträger ( $M = 1$ ) durchzuführen. Im allgemeinen regressionsanalytischen Modell können jedoch auch Zeitreihen mit voneinander abhängigen Kriteriumsvariablen (z.B. Korrelation von  $y_n$  mit  $y_{n+1}$ ) behandelt werden, weshalb die Prognose einer Serie von unbekannten Werten wünschenswert ist.

Für die Prognose muß gelten, daß das allgemeine regressionsanalytische Modell auch unter Hinzunahme von  $X_z$  und  $y_z$  noch gilt. Weil in den Annahmen A1 bis A5 diese zusätzlichen Variablen nicht berücksichtigt waren, müssen wir sie in Form eines *Prognosemodells* erweitern:

(P1) Es besteht die Beziehung

$$y_z = X_z \beta + u_z$$

mit demselben Vektor  $\beta$  wie in A1.

Dies bedeutet mit  $y^* = \begin{pmatrix} y \\ y_z \end{pmatrix}$ ,  $X^* = \begin{pmatrix} X \\ X_z \end{pmatrix}$  und  $u^* = \begin{pmatrix} u \\ u_z \end{pmatrix}$  insgesamt

$$y^* = X^* \beta + u^*$$

- (P2) Die Matrix  $X_z$  ist vom Rang  $M$ ,  $M \leq K+1$   
 (P3) Der Erwartungswert von  $u_z$  ist der Nullvektor:  $E\mathbf{u}_z = 0$   
 (P4) Die Kovarianzmatrix  $\Sigma_{u^*u^*}$  zu  $u^*$  existiert und ist nichtsingulär.

$$\text{Zerlegung: } \Sigma_{u^*u^*} = \sigma^2 \mathbf{V}^* = \sigma^2 \begin{pmatrix} \mathbf{V} & \mathbf{W} \\ \mathbf{W}' & \mathbf{V}_z \end{pmatrix}$$

- (P5) Der Vektor  $u_z$  ist  $M$ -dimensional normalverteilt.

Zuerst wollen wir uns der Prognose von

$$(24) \quad \mu_z = E\mathbf{y}_z,$$

also der *Prognose des Erwartungswertes* des Kriteriums für die zukünftigen Beobachtungen zuwenden. Die beste lineare erwartungstreue *Punktschätzung* dafür ist

$$(25) \quad \hat{\mu}_z = \mathbf{X}_z \hat{\beta}$$

mit  $\hat{\beta}$  aus Gleichung (17) (s. Schönfeld, 1969, p. 77), wobei nur die Annahmen A1 bis A4, P1 und P3 erforderlich sind. Sind alle Annahmen A1 bis A5 und P1 bis P5 erfüllt, ist eine *Bereichsschätzung* für  $\mu_z$  durch

$$(26) \quad \frac{1}{M\hat{\sigma}^2} (\mu_z - \hat{\mu}_z)' \mathbf{X}_z (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}'_z (\mu_z - \hat{\mu}_z) \leq F_\alpha(M, N-K-1)$$

mit der Irrtumswahrscheinlichkeit  $\alpha$  gegeben (Seifert, 1975, p. 87).

Bestehen Zusammenhänge zwischen den Fehlern  $u$  bei der ursprünglichen Stichprobe und den Fehlern  $u_z$  bei den hinzugekommenen Erhebungseinheiten, d.h. ist die Matrix  $W$  aus Annahme P4 keine Nullmatrix, dann kann man aus den geschätzten Fehlern  $\hat{u}$  sogar die Fehler  $u_z$  schätzen, so daß  $y_z = X \beta + u_z$  insgesamt genauer geschätzt werden kann, als dies wie in Gleichung (25) über eine alleinige Schätzung von  $\beta$  möglich ist. Insofern geht die *Prognose der Zufallsgröße*  $y_z$  über die Prognose des Erwartungswertes  $\mu_z$  hinaus. Sind die Annahmen A1 bis A4 und P1, P3 und P4 erfüllt, dann ist die beste (im Sinne des mittleren quadratischen Prognosefehlers) erwartungswert-erhaltende (d.h. der Erwartungswert der Schätzung  $\hat{y}_z$  ist gleich dem Erwartungswert von  $y_z$ ) lineare *Punktschätzung* für  $y_z$

$$(27) \quad \hat{y}_z = \mathbf{X}_z \hat{\beta} + \mathbf{W}' \mathbf{V}^{-1} (y - \mathbf{X} \hat{\beta}).$$

Für diese Punktschätzung muß außer der Matrix  $V$ , die wir grundsätzlich als bekannt vorausgesetzt haben, auch noch die Matrix  $W$  aus P4 bekannt sein. Wenn wie im klassischen regressionsanalytischen Modell  $W = 0$  gilt, fällt die Punktschätzung für  $y_z$  mit der für  $\mu_z$  zusammen.

Als Bereichsschätzung für  $y_z$  wünscht man sich eigentlich einen aufgrund der Realisation von  $y$  konstruierten Bereich, der dann mit der Wahrscheinlichkeit  $1 - \alpha$  den Zufallsvektor  $y_z$  enthält. Das bedeutet im „long run“, daß  $(1 - \alpha) \cdot 100\%$  der zu prognostizierenden Werte  $y_z$  in dem aufgrund der Realisierung von  $y$  konstruierten Bereich enthalten sind. Aufgrund rechnerischer Schwierigkeiten ist die Konstruktion solcher Bereiche im allgemeinen (wie auch im klassischen) regressionsanalytischen Modell noch nicht gelungen. Ein nicht voll befriedigender Lösungsansatz besteht in der Konstruktion von Toleranzintervallen und ist in Schneeweiß (1974, p. 81) skizziert. Wir wollen uns mit einer anderen Lösung begnügen; ihre Spezialisierung auf die klassische Regression mit nur einer Prädiktorvariablen führt zu den Vertrauensintervallen für Werte des Kriteriums, die in der elementaren Literatur (z.B. Hays, 1963, p. 523; Gaensslen & Schubö, 1976, p. 56; Bortz, 1977, p. 231) angegeben sind.

Wir stellen uns vor, daß einer Realisation von  $y$  auch nur eine Realisation von  $y_z$  zugeordnet wird. Die Bereichsschätzung wird so konstruiert, daß die Wahrscheinlichkeit dafür  $1 - \alpha$  ist, daß die Zufallsgröße  $y_z$  in dem durch  $y$  bestimmten Bereich liegt. Man stellt sich für die long run frequency-Interpretation von  $1 - \alpha$  also die wiederholte Realisation von  $y$  und  $y_z$  vor. Wenn die Annahmen A1 bis A5 und P1 bis P5 erfüllt sind, ist

$$(28) \quad \frac{1}{M\hat{\sigma}^2}(y_z - \hat{y}_z)' (\mathbf{V}_z + \tilde{\mathbf{X}}_z (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\tilde{\mathbf{X}}'_z - \mathbf{W}'\mathbf{V}^{-1}\mathbf{W})^{-1} (y_z - \hat{y}_z) \leqq F_{\alpha}(M, N-K-1)$$

mit  $\tilde{\mathbf{X}}_z = \mathbf{X}_z - \mathbf{W}'\mathbf{V}^{-1}\mathbf{X}$

eine Bereichsschätzung für  $y_z$  (s. Theil, 1971, p. 288) zur Irrtumswahrscheinlichkeit  $\alpha$ . Zur Berechnung dieser Bereichsschätzung ist die Kenntnis der ganzen Matrix  $\mathbf{V}^*$  aus P4 erforderlich, also die Kenntnis von  $V$ ,  $W$  und  $V_z$ .

Im klassischen regressionsanalytischen Modell (Unabhängigkeit aller Fehlervariablen aus  $u^*$ ) wird  $V = I$ ,  $W = 0$  und  $V_z = I$  vorausgesetzt, wodurch sich (28) zur Bereichsschätzung

$$(29) \quad \frac{1}{M\hat{\sigma}^2}(y_z - \hat{y}_z)' (\mathbf{I} + \mathbf{X}_z (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_z)^{-1} (y_z - \hat{y}_z) \leqq F_{\alpha}(M, N-K-1)$$

vereinfacht. Bei weiterer Vereinfachung auf eine Prädiktorvariable  $x$  ( $K = 1$ ) und eine zusätzliche Erhebungseinheit ( $M = 1$ ) erhält man die bekannte einfachste Bereichsschätzung für die skalare Zufallsvariable  $y_z$

$$(30) \quad \frac{(y_z - \hat{y}_z)^2}{\hat{\sigma}^2 \left( 1 + \frac{x_z^2 - 2x_z\bar{x} + \bar{x}^2}{N(\bar{x}^2 - \bar{x}^2)} \right)} \leqq F_{\alpha}(1, N-2).$$

## 1.5 Statistische Tests im klassischen regressionsanalytischen Modell

In den beiden letzten Kapiteln haben wir im Rahmen eines Modells über den funktionalen Zusammenhang versucht, Parameter oder von diesen abhängige Zufallsgrößen zu schätzen. Insofern war die Aufgabe die „Quantifizierung einer Theorie“. Jetzt wollen wir uns der Frage zuwenden, ob unsere Annahmen im Widerspruch zu den Beobachtungen stehen. Die neue Aufgabe ist also die „empirische Überprüfung einer Theorie“ (Seifert, 1975, p. 130). Dabei ist es häufig unmöglich, aus den Beobachtungen abzulesen, welche spezielle Annahme nicht zutrifft, weil der Widerspruch in der Regel nur innerhalb eines ganzen Satzes von Annahmen (Hypothesen und Oberhypothesen) hergestellt werden kann. Man muß dann nach dem Ausschlußverfahren der Reihe nach alle Annahmen bis auf eine von ihnen als Quelle des Widerspruchs ausschließen.

Eine Übersicht über die Theorie der statistischen Tests wird in diesem Handbuchband von Wendt gegeben. Doch wollen wir hier die Nomenklatur der statistischen Tests (s. auch Haagen & Seifert, 1979, Teil III) präzisieren. Unter einem *Signifikanztest* verstehen wir die Konstruktion eines kritischen Bereichs derart, daß für alle Parameter aus der Nullhypothese die Wahrscheinlichkeit dafür  $a$  ist, daß die Prüfgröße in den kritischen Bereich fällt. Die Entscheidungsregel ist: Fällt die Realisation der Prüfgröße in den kritischen Bereich, verwerfen wir die Nullhypothese; sonst enthalten wir uns des Urteils. Im Neyman-Pearson-Test (NP-Test) wird der kritische Bereich so konstruiert, daß die Wahrscheinlichkeit dafür, daß die Prüfgröße für alle Parameter aus der Nullhypothese einen Wert in diesem Bereich annimmt,  $a$  ist, und möglichst so, daß sie für alle Parameter aus der Alternativhypothese maximal ist; gelingt dies, nennt man den Test einen gleichmäßig mächtigsten Test (UMP-Test). Bei sog. zweiseitigen Alternativhypothesen gibt es keinen gleichmäßig mächtigsten Test; wenn man vom kritischen Bereich jedoch fordert, daß die Wahrscheinlichkeit dafür, daß die Prüfgröße für die Parameter der Alternativhypothese einen Wert in diesem Bereich annimmt, stets mindestens  $a$  ist, existiert auch dann ein gleichmäßig mächtigster Test, den man einen gleichmäßig mächtigsten unverfälschten Test (UMPU-Test) nennt. Beim *Likelihood-Quotienten-Test* (LQ-Test) wird als Prüfgröße der Likelihood-Quotient verwendet, der Quotient der maximalen Likelihood aus der Nullhypothese und der maximalen Likelihood aus der Alternativhypothese. Als kritischer Bereich wird ein Intervall zwischen 0 und dem kritischen LQ gewählt. Sowohl beim NP-Test als auch beim LQ-Test entscheidet man sich für die Nullhypothese, wenn die Prüfgröße nicht in den kritischen Bereich fällt.

Aus Bereichsschätzungen zu einem evtl. vektoriellen Parameter  $\Theta$  kann immer zumindest ein Signifikanztest konstruiert werden, wenn die Nullhypothese im

zu schätzenden Parameter einfach ist. Hat die Bereichsschätzung zum Signifikanzniveau  $\alpha$  die Gestalt  $f(\Theta, y) \leq g_\alpha$  und spezifiziert die Nullhypothese den Parameter  $\Theta = \Theta^{(0)}$ , dann weiß man, daß  $f(\Theta^{(0)}, y) \leq g_\alpha$  mit Wahrscheinlichkeit  $1 - \alpha$  gilt, wenn die Nullhypothese zutrifft. Als Prüfgröße kann daher  $f(\Theta^{(0)}, y)$  verwendet werden; ist ihre Realisation  $f(\Theta^{(0)}, y) > g$ , entscheidet man sich gegen  $\Theta^{(0)}$  und damit gegen die Nullhypothese. Da die Tests für das allgemeine regressionsanalytische Modell durch die beschriebenen Bereichsschätzungen bereits explizit gegeben sind, werden hier die entsprechenden Tests für das *klassische regressionsanalytische Modell* mit Normalverteilungsannahme A5 angegeben.

Zur Überprüfung von *Hypothesen über die Regressionskoeffizienten* verwenden wir wieder die Linearkombination  $\gamma = L\beta$  aus Gleichung (20) mit der  $M \times (K+1)$ -Kontrastmatrix  $L$ . Die Nullhypothese sei  $\gamma = \gamma^{(0)}$ , die Alternativhypothese  $\gamma \neq \gamma^{(0)}$ . Die Nullhypothese wird verworfen, wenn

$$(31) \quad \frac{1}{M\hat{\sigma}^2} (\gamma^{(0)} - \hat{\gamma})' (L(X'X)^{-1}L')^{-1} (\gamma^{(0)} - \hat{\gamma}) > F_{\alpha}(M, N-K-1)$$

mit  $\hat{\gamma} = L(X'X)^{-1}X'y$

$$\text{und } \hat{\sigma}^2 = \frac{1}{N-K-1} y' (I - X (X'X)^{-1}X')y$$

ist. Dies ist auch gleichwertig dem LQ-Test und für  $M = 1$  (nur eine skalare Linearkombination) ein gleichmäßig mächtigster unverfälschter NP-Test, wenn in der Entscheidungsregel die Entscheidung  $H_0$  bei „ $\leq$ “ in Gleichung (31) einbezogen wird. Für  $M > 1$  existiert kein gleichmäßig mächtigster unverfälschter NP-Test (Kendall & Stuart, 1976, p. 238). Wenn als Kontrastmatrix  $L$  die Einheitsmatrix verwendet wird, kommt man zum Test der Regressionskoeffizienten. Bei der speziellen Nullhypothese  $\beta^{(0)} = 0$  erhält man mit dem aus Gleichung (15) übernommenen Bestimmtheitsmaß (Quadrat der empirischen multiplen Korrelation)

$$(32) \quad R^2 = \frac{s_y^2}{s_y^2} = \frac{\hat{y}' \hat{y} - N\bar{y}^2}{y'y - N\bar{y}^2} = \frac{y'X(X'X)^{-1}X'y - N\bar{y}^2}{y'y - N\bar{y}^2}$$

die Entscheidungsregel, daß die Nullhypothese verworfen wird, wenn

$$(33) \quad \frac{(N-K-1) \left( R^2 + \frac{\bar{y}^2}{s_y^2} \right)}{(K+1)(1-R^2)} > F_{\alpha}(K+1, N-K-1).$$

Dieser Test wird in der Regel nur sinnvoll sein, wenn sowohl die Prädiktorenvariablen als auch die Kriteriumsvariable auf Verhältnisskalen (z.B. Gigerenzer, 1981, p. 116) gemessen sind, weil nur dann die Komponente  $\beta_0$  des Vektors

(der Achsenabschnitt der Regressionsfläche) in allen zulässigen Skalierungen der Variablen zugleich den Wert Null hat. Beschränkt man die Nullhypothese auf die „echten“ Regressionskoeffizienten  $\beta_1^{(0)} = \beta_2^{(0)} = \dots = \beta_K^{(0)} = 0$  und läßt das Absolutglied  $\beta_0^{(0)}$  (wie stets auch  $\sigma^2$ ) offen, erhält man aus Gleichung (31) mit einer geeigneten Matrix L die bekannte Entscheidungsregel zum Verwerfen der Nullhypothese:

$$(34) \quad \frac{(N-K-1) R^2}{K (1-R^2)} > F_\alpha(K, N-K-1)$$

(Cooley & Lohnes, 1971, p. 51; Kerlinger & Pednazar, 1973, p. 63; Cohen & Cohen, 1975, p. 104; Gaensslen & Schubö, 1976, p. 108; Bortz, 1977, p. 592; Moosbrugger, 1978, p. 70; Küchler, 1979, p. 127). Mit diesem Test kann also über die Frage entschieden werden, ob wenigstens eine der Prädiktorvariablen  $x_1, \dots, x_K$  einen von Null verschiedenen Regressionskoeffizienten hat, ob also insgesamt ein Zusammenhang zwischen Kriterium und Prädiktoren besteht.

Bei der Prüfung der Frage nach der *Relevanz einzelner Prädiktorvariablen*  $x_k$  ist der schwerwiegendere Fehler, wenn eine bedeutsame Variable vergessen wird, als wenn eine Variable mit  $\beta_k = 0$  überflüssigerweise aufgenommen wird. Denn das Vergessen einer bedeutsamen Variablen ist ein Spezifikationsfehler in A1, der zur Folge haben kann, daß die Schätzung der Regressionskoeffizienten zu den übrigen Variablen nicht erwartungstreu ist, sowie daß sonst vermeidbare Heteroskedastizität, Autokorrelation oder Strukturbruch auftreten. Daher wäre der Signifikanztest der Nullhypothese  $\beta_k^{(0)} \neq 0$  angemessen; ein derartiger Signifikanztest bereitet jedoch erhebliche theoretische Schwierigkeiten und kann höchstens für feste Werte, z.B.  $\beta_k^{(0)} = -0.5$ , durchgeführt werden. Man begnügt sich in der Regel mit der Prüfung der Nullhypothese  $\beta_k^{(0)} = 0$ , wodurch nur das Signifikanzniveau  $\alpha$  für den weniger gravierenden Fehler kontrolliert wird. Entsprechend der Bereichsschätzung (23) kommt man zur Entscheidungsregel, daß die Nullhypothese abgelehnt wird, wenn

$$(35) \quad \frac{b_k^2}{\hat{\sigma}_{b_k}^2} > F_\alpha(1, N-K-1)$$

ist. Auch dieser Test kann durch entsprechende Korrelationskoeffizienten ausgedrückt werden. Dazu muß aber noch die empirische semipartiale Korrelation (vgl. Gaensslen & Schubö, 1976, p. 87)

$$(36) \quad r_{y_{(k-12 \dots k \dots K)}} = b_k \sqrt{\frac{1 - R^2}{\hat{\sigma}_{b_k}^2(N-K-1)}}$$

eingeführt werden. Man erhält dann dieselbe Entscheidungsregel wie bei (35) mit

$$(37) \quad \frac{(N-K-1)r_{y_{(k-12 \dots k \dots K)}}^2}{1 - R^2} > F_\alpha(1, N-K-1).$$

Weil hier nur ein Regressionskoeffizient geprüft wird, ist der Test als NP-Test ein gleichmäßig mächtigster unverfälschter Test. Einen gleichmäßig mächtigsten Test für die einseitige Alternativhypothese  $\beta_k > 0$  erhält man durch

$$\frac{b_k}{\sqrt{\hat{\sigma}_{b_k}^2}} > \sqrt{F_{2\alpha}(1, N-K-1)} = t_\alpha(N-K-1).$$

Als Beispiel für die Anwendung einer Kontrastmatrix soll der Test auf Strukturbruch vorgestellt werden. Unter Strukturbruch versteht man eine Verletzung des klassischen regressionsanalytischen Modells derart, daß entweder die Regressionskoeffizienten  $\beta$  oder die Fehlervarianz  $\sigma^2$  zu den ersten  $N_1$  Beobachtungen einen anderen Wert haben als die zu den zweiten  $N_2 = N - N_1$  Beobachtungen. Das Problem und dessen Behandlung werden unter dem Stichwort „*differentialle Prognostizierbarkeit*“ angesprochen (z.B. Moosbrugger, 1980b). Wenn der Strukturbruch erkannt ist, müssen entweder in den beiden Gruppen getrennte Modelle aufgestellt werden oder man muß das Regressionsmodell um sog. Interaktionen erweitern. Für den Test auf Strukturbruch der Fehlervarianz werden wir das erste Vorgehen wählen, das zweite Vorgehen wählen wir für den Test auf Strukturbruch in den Regressionskoeffizienten. Hier soll zunächst der *Strukturbruch in den Regressionskoeffizienten* getestet werden. Für die Darstellung beschränken wir uns auf die Einfachregression  $K = 1$  mit  $\beta' = (\beta_0, \beta_1)$ .  $\beta_0$  ist der Koeffizient zur Konstante  $x_{0n} = 1$ ,  $\beta_1$  der zu  $x_{1n} = x_n$ .

Die Regressionskoeffizienten in der 1. Beobachtungsgruppe seien  $\beta_0^{(1)}, \beta_1^{(1)}$ , die in der zweiten Beobachtungsgruppe  $\beta_0^{(2)}, \beta_1^{(2)}$ . Die Fehlervarianz sei in beiden Gruppen einheitlich  $\sigma^2$ . Um den Test (31) anwenden zu können, muß das Modell auf 4 Regressionskoeffizienten, d.h. auch auf 4 Prädiktoren erweitert werden:

$$(38) \quad y_n = \beta_0^{(1)} \delta_n^{(1)} + \beta_1^{(1)} \delta_n^{(1)} x_n + \beta_0^{(2)} \delta_n^{(2)} + \beta_1^{(2)} \delta_n^{(2)} x_n + u_n$$

mit den „Dummyvariablen“

$$(39) \quad \delta_n^{(i)} = \begin{cases} 1 & \text{falls Beobachtung } n \text{ zur Gruppe } i \text{ gehört} \\ 0 & \text{sonst.} \end{cases}$$

Dadurch, daß nun insgesamt 4 Prädiktoren vorliegen, von denen für jede Gruppe jeweils zwei konstant den Wert Null haben, liegen in einer Regressionsgleichung zwei verschiedene Regressionsansätze vor. Obwohl wegen des schwerwiegenderen Fehlers, den Strukturbruch zu übersehen, eigentlich die gegensätzliche Hypothesenstellung angemessen wäre, wählt man wegen der erwähnten Schwierigkeiten als Nullhypothese

$$(40) \quad \beta_0^{(1)} = \beta_0^{(2)} \text{ und } \beta_1^{(1)} = \beta_1^{(2)}$$

oder  $\gamma = \begin{pmatrix} \beta_0^{(1)} - \beta_0^{(2)} \\ \beta_1^{(1)} - \beta_1^{(2)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$

Die Kontrastmatrix L ist demnach

$$(41) \quad L = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix}$$

womit die Prüfung nach (31) erfolgen kann. Praktisch wird dies nur möglich sein, wenn ein Computerprogramm zum Linearen Modell vorliegt. Für Computerprogramme zur Regressionsanalyse ist der Test für den zusätzlichen Beitrag von Variablengruppen nach Kap. 1.11 für den folgenden gleichwertigen Modellansatz günstiger

$$(42) \quad Y_n = \beta_0 + \beta_1 x_n + \beta_2 (\delta_n^{(1)} - \delta_n^{(2)}) + \beta_3 (\delta_n^{(1)} - \delta_n^{(2)}) x_n + u_n,$$

wobei der Zusammenhang zu den Regressionskoeffizienten aus dem Ansatz (38) durch

$$(43) \quad \begin{aligned} \beta_0 &= (\beta_0^{(1)} + \beta_0^{(2)})/2 \\ \beta_1 &= (\beta_1^{(1)} + \beta_1^{(2)})/2 \\ \beta_2 &= (\beta_2^{(1)} - \beta_2^{(2)})/2 \\ \beta_3 &= (\beta_3^{(1)} - \beta_3^{(2)})/2 \end{aligned}$$

gegeben ist. Die Nullhypothese (40) lautet in dieser Parametrisierung jetzt einfacher

$$(44) \quad \beta_2 = \beta_3 = 0$$

und kann als Test direkt zu den Regressionskoeffizienten praktisch leichter durchgeführt werden. Eine derartige Reparametrisierung ist bei jeder Kontrastmatrix L möglich, wobei lediglich die Prädiktorvariablen entsprechend transformiert werden müssen.

Der Test auf *Strukturbruch der Fehlervarianz* prüft, ob die Fehlervarianzen  $\sigma_{(1)}^2$  und  $\sigma_{(2)}^2$  zu den beiden Beobachtungsgruppen gleich sind. Insofern handelt es sich um eine einfachste Form der Prüfung auf *Varianzheterogenität* (Heteroskedastizität); differenziertere, kompliziertere und auch nur asymptotisch gültige Tests hierzu sind z.B. in Ramsey (1969), Kmenta (1971, p. 269) und Gartside (1972) beschrieben. Wie angekündigt stellen wir zwei getrennte Regressionsansätze in den beiden Beobachtungsgruppen auf; dadurch wird auch zugelassen, daß zugleich Strukturbruch in den Regressionskoeffizienten besteht. Aufgrund der beiden Teilstichproben vom Umfang  $N_1$  und  $N_2$  bildet man Schätzungen der Fehlervarianzen  $\hat{\sigma}_{(1)}^2$  und  $\hat{\sigma}_{(2)}^2$ . Für die einseitige Fragestellung mit der Alternativhypothese  $\sigma_{(1)}^2 < \sigma_{(2)}^2$  erhält man mit der Entscheidungsregel

$$(45) \quad \frac{\hat{\sigma}_{(1)}^2}{\hat{\sigma}_{(2)}^2} < F_{1-\alpha}(N_1-K-1, N_2-K-1) = \frac{1}{F_\alpha(N_2-K-1, N_1-K-1)}$$

für die Alternativhypothese einen gleichmäßig mächtigsten NP-Test zum Niveau  $\alpha$ . Bei ungerichteter Alternativhypothese  $\sigma_{(1)}^2 \neq \sigma_{(2)}^2$  ist wegen der Asymmetrie der F-Verteilung folgender zweiseitiger Test mit jeweils gleichen Irrtumswahrscheinlichkeiten ein nur näherungsweise gleichmäßig mächtigster unverfälschter NP-Test: Verwirf die Nullhypothese zugunsten der Alternativhypothese, wenn

$$(46) \quad \frac{\hat{\sigma}_{(1)}^2}{\hat{\sigma}_{(2)}^2} < F_{1-\alpha/2}(N_1-K-1, N_2-K-1) \quad \text{oder}$$

$$\frac{\hat{\sigma}_{(1)}^2}{\hat{\sigma}_{(2)}^2} > F_{\alpha/2}(N_1-K-1, N_2-K-1).$$

Besonders schwierig zu interpretieren sind *Prognosetests*, weil durch sie die Nullhypothese „der ganze Satz von Annahmen A1 bis A5 und P1 bis P5 trifft zu“ geprüft wird. Wird durch den Test die Nullhypothese verworfen, muß durch das Vorwissen eingegrenzt werden, welche der Annahmen speziell fragwürdig ist, ob etwa ein Strukturbruch bei der Erhebung der zusätzlichen Beobachtungseinheiten vorliegt, ob ein Modellfehler in Form eines vergessenen Prädiktors vorliegt oder die Normverteilungsannahme verletzt ist (oder im allgemeinen regressionsanalytischen Modell die Matrix V falsch spezifiziert wurde). Dennoch hat der Prognosetest für die Beurteilung einer Regressionsbeziehung als ganzes seine unbestrittene Bedeutung; er ist für *Kreuzvalidierungsstudien* der Test der Wahl. Weil die Anwendung der Bereichsschätzung (28) für den Prognosetest im allgemeinen regressionsanalytischen Modell allenfalls rechnerische Schwierigkeiten mit sich bringt, wird auch hier wie überall in diesem Kapitel 1.5 nur die Spezialisierung auf das klassische regressionsanalytische Modell dargestellt. Gegenüber der Anwendung der Bereichsschätzung (29) als Signifikanztest kann ein etwas vereinfachter Test dadurch erreicht werden, daß die Prädiktoren in der Matrix X „ortho-standardisiert“ werden; dies kann z.B. durch eine Hauptkomponentenanalyse mit Bildung von standardisierten Hauptkomponentenwerten geschehen, wobei die Konstante  $x_0 = 1$  nicht transformiert werden muß. Für derartige Prädiktoren gilt

$$(47) \quad X'X = NI,$$

womit der Prognosetest durch die Entscheidungsregel zur Verwerfung der Nullhypothese

$$(48) \quad \frac{1}{M\hat{\sigma}^2} (y_z - \hat{y}_z)' (I + \frac{1}{N} X_z X_z')^{-1} (y_z - \hat{y}_z) > F_\alpha(M, N-K-1)$$

gegeben ist. Dabei ist  $\hat{\sigma}^2$  die Schätzung der Fehlervarianz aus der ursprünglichen Stichprobe und  $M$  die Anzahl der Beobachtungseinheiten in der zusätzlichen Stichprobe. Die Schwierigkeit bei dieser Entscheidungsregel ist, daß sie mit den üblichen Computerprogramm nicht auszuwerten ist. Doch kann man

für die linke Seite der Ungleichung (48) eine obere und untere Schranke angeben, deren Komponenten z.B. mit dem Statistik-Programmsystem SPSS (Nie et al., 1975; Beutel et al., 1980) bestimmt werden können. Wenn (47) nach wie vor erfüllt ist, gilt mit

$$(49) \quad h_{ob} = \sum_{n=N+1}^{N+M} (y_n - \hat{y}_n)^2 \quad \text{und}$$

$$h_{unt} = \sum_{n=N+1}^{N+M} (y_n - \hat{y}_n)^2 - \frac{1}{N} \sum_{k=0}^K \left( \sum_{n=N+1}^{N+M} (y_n - \hat{y}_n) x_{nk} \right)^2$$

die Ungleichung

$$h_{unt} \leq (y_z - \hat{y}_z)' (I + \frac{1}{N} X_z X_z')^{-1} (y_z - \hat{y}_z) \leq h_{ob}.$$

Mit diesen Größen kommen wir zur Entscheidungsregel:

Für

$$\frac{h_{unt}}{M\hat{\sigma}^2} > F_{\alpha}(M, N-K-1): \text{Entscheidung für } H_1$$

(50)

$$\frac{h_{ob}}{M\hat{\sigma}^2} < F_{\alpha}(M, N-K-1): \text{Entscheidung für } H_0$$

sonst: keine Entscheidung oder Rückgriff auf (48).

Der *Durbin-Watson-Test* auf Verletzung der Annahme des klassischen regressionsanalytischen Modells, daß die Fehlervariablen unkorreliert sind, wird in Kapitel 1.9 besprochen. Die *Normalverteilungsannahme* verbleibt schließlich als eine schwer überprüfbare Annahme. Im Modell-Ansatz verlangt A5, daß die Fehler  $u_n$  multivariat normalverteilt sind. Dafür genügt es nicht, daß jede Fehlervariable  $u_n$  für sich normalverteilt ist, wenn nicht zusätzlich über K4 hinaus die Unabhängigkeit dieser Fehlervariablen angenommen werden kann. Es ist wegen des zentralen Grenzwertansatzes richtig, daß die Fehler dann mit guter Näherung als multivariat normalverteilt betrachtet werden können, wenn sie additiv auf viele, unabhängige Einflußfaktoren zurückgehen. Doch niemand kann letztlich ausschließen, daß der Fehler fast ausschließlich auf eine im Modell nicht berücksichtigte, nicht normal verteilte Variable zurückgeht. Draper & Smith (1966, p. 87) diskutieren einige Möglichkeiten der grafischen Beurteilung der Normalverteilungsannahme, die sich auf die empirische Verteilung der N geschätzten Residuen stützen. Eine einfache Methode zur Beurteilung der univariaten Normalität, die vor allem auf Ausreißer in der empiri-

schen Verteilung der Residuen empfindlich ist, besteht darin, nachzuprüfen, ob im Bereich

$$[-\sqrt{\hat{\sigma}^2} t_{\alpha/2}(N-K-1), +\sqrt{\hat{\sigma}^2} t_{\alpha/2}(N-K-1)]$$

auch wirklich der Anteil  $1 - \alpha$  der Residuen liegt. In Tests zur Normalverteilungsannahme führen Lienert (1967, p. 172) und Koerts & Abrahamse (1969, Kap. 7) ein. Über die Robustheit der Bereichsschätzungen und Signifikanztests gegenüber Abweichungen von der Normalverteilung findet man in Malinvaud (1966, p. 297) Hinweise.

## 1.6 Ridge-Regression

Wir betrachten das klassische Regressionsmodell  $y = X\beta + u$ , bei dem die Annahmen A1, A2, A3 und K4 erfüllt sind, so daß die Fehler  $u_k$  unkorreliert sind und gleiche Varianz haben. Es sollen nun genauer die Eigenschaften des Minimum-Quadrat-Schätzers

$$\mathbf{b} = (X'X)^{-1}X'y$$

untersucht werden. Im Zusammenhang mit Gl. (18) war berichtet worden, daß  $\mathbf{b}$  der beste lineare erwartungstreue Schätzer ist. Diese Aussage des Satzes von Gauß-Markov bedeutet, daß für alle Linearkombinationen  $\mathbf{l}'\mathbf{b}$  mit festem, beliebigem Gewichtvektor  $\mathbf{l}$  die Varianz kleiner ist als für die mit irgendeinem anderen linearen erwartungstreuen Schätzer  $\check{\beta}$  gebildete Linearkombination  $\mathbf{l}'\check{\beta}$ :

$$\text{var}(\mathbf{l}'\mathbf{b}) \leq \text{var}(\mathbf{l}'\check{\beta})$$

für beliebige  $\mathbf{l}$  und alle linearen erwartungstreuen Schätzer  $\check{\beta}$ . Setzt man für  $\mathbf{l}$  den Einsvektor  $\mathbf{j}$  ein, erhält man als Varianz

$$(51) \quad \text{var}(\mathbf{j}'\mathbf{b}) = E(\mathbf{b} - \beta)'(\mathbf{b} - \beta) = E \sum_{k=0}^K (b_k - \beta_k)^2$$

den Erwartungswert der Summe der quadrierten Schätzfehler, den man auch den mittleren quadratischen Fehler (mean square error) nennt. Dieser mittlere quadratische Fehler ist für den Minimum-Quadrat-Schätzer  $\mathbf{b}$  also minimal unter den linearen und erwartungstreuen Schätzern. Dieser Fehler wird sehr groß, wenn die Prädiktoren angenähert kollinear sind, wie sich aus den folgenden Überlegungen ergibt.

Für die Kovarianzmatrix von  $\mathbf{b}$  ergibt sich

$$(52) \quad \Sigma_{bb} = \sigma^2 (X'X)^{-1}.$$

Nach Definition der Kovarianzmatrix gilt

$$(53) \quad E \mathbf{b} \mathbf{b}' = \boldsymbol{\beta} \boldsymbol{\beta}' + \Sigma_{bb}$$

und damit gemäß (52)

$$(54) \quad E \mathbf{b} \mathbf{b}' = \boldsymbol{\beta} \boldsymbol{\beta}' + \sigma^2 (\mathbf{X}' \mathbf{X})^{-1}.$$

Daraus ergibt sich weiter

$$(55) \quad E \sum_k b_k^2 = \text{Spur}(E \mathbf{b} \mathbf{b}') = \sum_k \beta_k^2 + \sigma^2 \text{Spur}(\mathbf{X}' \mathbf{X})^{-1}.$$

Bezeichnen wir mit  $\lambda_1, \dots, \lambda_K$  die Eigenwerte von  $\mathbf{X}' \mathbf{X}$ , dann ergibt sich aus (55)

$$(56) \quad E \sum_k b_k^2 = \sum_k \beta_k^2 + \sigma^2 \sum_k (1/\lambda_k).$$

Aus (56) ist ersichtlich, daß der Erwartungswert der Summe der Quadrate der  $b_k$  um so mehr von der Summe der Quadrate der  $\beta_k$  abweicht, je kleinere Eigenwerte von  $\mathbf{X}' \mathbf{X}$  auftreten. Wenn aber die Prädiktoren fast kollinear sind, so ist mindestens ein Eigenwert von  $\mathbf{X}' \mathbf{X}$  relativ klein. Dementsprechend wird auch der mittlere quadratische Fehler

$$(57) \quad E \sum_k (b_k - \beta_k)^2 = \sigma^2 \sum_k (1/\lambda_k)$$

relativ groß, so daß sich im Fall angenäherter Kollinearität sehr ungenaue Schätzungen ergeben. Die Ungenauigkeit der Schätzungen beruht letztlich auf mangelnder Information und könnte nur durch zusätzliche a priori Information kompensiert werden.

Um nun zu verhindern, daß die erwartete Quadratsumme der Schätzungen zu groß wird (siehe (56)), werden die  $\beta_k$  nach der Minimum-Quadrat-Methode durch einen Vektor  $\tilde{\beta}$  unter der folgenden Nebenbedingung

$$(58) \quad \sum_k \tilde{\beta}_k^2 = c$$

geschätzt. Mit Hilfe des Lagrange-Ansatzes ergibt sich das Minimierungsproblem

$$(59) \quad (\mathbf{y} - \mathbf{X} \tilde{\beta})' (\mathbf{y} - \mathbf{X} \tilde{\beta}) + h \tilde{\beta}' \tilde{\beta} \stackrel{!}{=} \text{Minimum},$$

wobei  $h$  der Lagrange-Multiplikator ist. Daraus resultieren die Normalgleichungen

$$(60) \quad \mathbf{X}' \mathbf{X} \tilde{\beta} + h \tilde{\beta} = \mathbf{X}' \mathbf{y}$$

bzw. die für  $h \geq 0$  stets existierende Lösung

$$(61) \quad \tilde{\beta} = (\mathbf{X}' \mathbf{X} + h \mathbf{I})^{-1} \mathbf{X}' \mathbf{y}.$$

Man könnte nun wegen (58)  $h$  durch  $c$  ausdrücken und als Funktion von  $c$  ersetzen. Da dieser Ersetzungsprozeß nur implizit möglich ist, parametrisieren wir  $\beta$  durch  $h$ , d.h. für jedes  $h$  wird vermöge (61) ein Schätzer festgelegt. Speziell resultiert für  $h = 0$  der gewöhnliche Minimum-Quadrat-Schätzer. Der Schätzer  $\beta$  heißt Ridge-Regression-Schätzer (R.R.-Schätzer).

Es gibt noch eine Verallgemeinerung, die sogenannte generalisierte Ridge-Regression-Schätzung (G.R.R.-Schätzer), bei der das Vielfache der Einheitsmatrix in (61) durch eine Diagonalmatrix  $H$  ersetzt wird

$$(62) \quad \tilde{\beta} = (\mathbf{X}'\mathbf{X} + H)^{-1}\mathbf{X}'\mathbf{y}.$$

Im folgenden wird nur die R.R.-Schätzung betrachtet. Zuerst wird gezeigt, daß der R.R.-Schätzer verzerrt ist. Es gilt nämlich

$$(63) \quad E(\tilde{\beta}) = (\mathbf{X}'\mathbf{X} + hI)^{-1}\mathbf{X}'(\mathbf{X}\beta) = \beta - h(\mathbf{X}'\mathbf{X} + hI)^{-1}\beta.$$

Also ergibt sich die Verzerrung

$$(64) \quad E(\tilde{\beta}) - \beta = -h(\mathbf{X}'\mathbf{X} + hI)^{-1}\beta.$$

die von  $\beta$  abhängt.

Der Vorteil des R.R.-Schätzers ist, daß es ein  $h$  gibt, so daß der mittlere quadratische Fehler  $E(\tilde{\beta} - \beta)'(\tilde{\beta} - \beta)$  kleiner ist als der des gewöhnlichen Minimum-Quadrat-Schätzers  $E(\mathbf{b} - \beta)'(\mathbf{b} - \beta)$ , wie Hoerl & Kennard (1970) gezeigt haben. Von Theobald (1974) wurde gezeigt, daß für

$$\langle h \rangle < h_{\max} = \frac{2\sigma^2}{\beta' \beta} \text{ und alle positiv definiten Matrizen } W \text{ gilt}$$

$$(65) \quad E(\tilde{\beta} - \beta)'W(\tilde{\beta} - \beta) < E(\mathbf{b} - \beta)'W(\mathbf{b} - \beta).$$

Wird speziell  $W = I$  gewählt, so ergibt sich eine Aussage, die der von Hoerl & Kennard (1970) entspricht. Da aber  $h_{\max}$  von den unbekannten  $\beta_k$  abhängt, hat die Aussage (65) nur geringe praktische Bedeutung.

Es gibt auch Versuche,  $h_{\max}$  und damit  $h$ , in Abhängigkeit von  $y$  und  $X$  zu schätzen. Dann werden  $h_{\max}$  und  $h$  stochastisch und der entsprechende R.R.-Schätzer nichtlinear. Die theoretischen Eigenschaften dieses Schätzers sind unbekannt.

Um sich einen Überblick über das Verhalten des R.R.-Schätzers in Abhängigkeit von  $h$  zu verschaffen, werden die  $\beta_k$  als Funktionen von  $h$  grafisch dargestellt. Diese Darstellung wird als Ridge Trace bezeichnet.

Für die praktische Anwendung des R.R.-Schätzers ist es von Bedeutung, daß die Minimum-Quadrat-Schätzung in folgendem Sinne instabil ist: Bei angenä-

hert kollinearren Prädiktoren ändern sich die  $b_k$  aufgrund geringerer Änderungen der Daten in erheblichem Maße. Für die R.R.-Schätzung verringert sich diese Instabilität; obwohl (und weil) sie nicht erwartungstreu ist, hat sie einen kleineren mittleren quadratischen Fehler.

Als ergänzende Lektüre können Hoerl & Kennard (1970), Theobald (1974), Vinod (1978) und Jahn (1979) empfohlen werden.

## 1.7 Klassisches korrelationsanalytisches Modell und multiple Korrelation

Bei der Diskussion des regressionsanalytischen Modells wurde auf die besondere Bedeutung der Annahme hingewiesen, daß die Werte X der Prädiktoren feste Werte sind. Dies bedeutet für die Interpretation der Wahrscheinlichkeit als relative Häufigkeit im „long run“, daß die Replikation der Beobachtungen zu immer denselben Prädiktorwerten erfolgen muß. Mit Replikation der Beobachtungen ist allerdings nicht die Wiederholung der Beobachtungen zum Aufbau einer Stichprobe, sondern die Wiederholung der Untersuchung an einer neuen Stichprobe gemeint. Die Annahme fester Prädiktoren ist für viele psychologische Fragestellungen unrealistisch. In diesem Kapitel soll gezeigt werden, unter welchen veränderten Annahmen mit etwas abweichender Interpretation die bisher berichteten Ergebnisse beibehalten werden können, wenn stochastische Prädiktoren vorliegen, also die Prädiktoren selbst Zufallsvariablen sind. Als nicht leicht verständliche, aber umfassende Lektüre zu diesem Thema können Schönfeld (1971, p.1) oder Graybill (1976, p. 373) empfohlen werden. Eine erste Einführung in das korrelationsanalytische Modell bieten Gaensslen & Schubö (1976, p. 47).

Das *klassische korrelationsanalytische Modell (general linear regression model, unabhängige stochastische lineare Regression)* umfaßt folgende Annahmen:

- (S1) Es besteht die Beziehung  

$$y = X\beta + u,$$
 wobei X eine  $N \times (K+1)$ -Matrix von beobachtbaren Zufallsgrößen,  $\beta$  ein Vektor von festen, unbekannten Parametern mit  $K+1$  Komponenten, y ein Vektor von N beobachtbaren Zufallsgrößen und u ein Vektor von N unbeobachtbaren Zufallsgrößen ist. X und u (und damit y) besitzen eine gemeinsame Wahrscheinlichkeitsverteilung.
- (S2) Der Rang der Zufallsmatrix X ist mit Wahrscheinlichkeit 1 gleich der Spaltenzahl  $K+1$ .
- (S3)  $E_u(u|X) = 0$  für alle Realisationen X.
- (S4)  $\Sigma(u,u|X) = E_u(u'u'|X) = \sigma^2 I$  für alle Realisationen X.

- (S5) Für jede Realisation  $X$  ist die bedingte Verteilung von  $u$  N-dimensional normal.
- (S6) Es existieren  $E(X'X)$ ,  $E(X'X)^{-1}$  und  $E(X'X)^{-1}X'$ .  
 $EX'X$  ist nichtsingulär.
- (S7) Die Rand-Dichtefunktion  $f_x(X)$  ist für alle Realisationen  $X$  positiv.

Die Annahmen S1 bis S5 entsprechen weitgehend den früheren Annahmen A1 bis A5. Um in nicht zu große Schwierigkeiten bei der Darstellung zu gelangen, wurde für S4 auf die Möglichkeit einer beliebigen Kovarianzmatrix verzichtet. Ansonsten wurde dem Umstand Rechnung getragen, daß nunmehr Zufallsvariablen auch für  $X$  vorliegen, wodurch auch die zusätzlichen Annahmen S6 und S7 erforderlich werden. Weiterhin wird vorausgesetzt, daß die Prädiktoren fehlerfrei gemessen sind.

Wichtig ist die Unterscheidung zwischen dem bedingten und unbedingten Erwartungswert; der bedingte Erwartungswert  $E_u(u|X)$  bedeutet den Erwartungswert von  $u$  unter der Bedingung, daß  $X$  einen festen Wert annimmt, der unbedingte Erwartungswert  $E_u(u) = Eu$  ist der Erwartungswert in der Randverteilung von  $u$ . Unter den Annahmen S3 und S7 gilt jedoch auch

$$(66) \quad E_{X,u}(u) = E_X(E_u(u|X)) = E_X(0) = 0,$$

so daß der unbedingte Erwartungswert wie der bedingte verschwindet. Ebenso ist wegen S4 und S7

$$(67) \quad \Sigma_{uu} = \sigma^2 I.$$

Durch Bildung des bedingten Erwartungswerts für die Regressionsbeziehung in S1 erhält man für jede Realisierung  $X$

$$(68) \quad E_u(y|X) = X\beta.$$

Insofern entspricht  $E_u(y|X)$  einerseits  $E_y$  aus Gleichung (16) und andererseits der Approximation  $\hat{y}$  aus Gleichung (4) der beschreibenden linearen Regression (vgl. Steyer, 1979; Moosbrugger, 1980a, 1980b). Als Konsequenzen der Annahmen des stochastischen Modells (ohne die Normalverteilungsannahme S5) gilt: Die Fehler  $u$  und die Prädiktoren  $X$  sind unkorreliert, können aber stochastisch abhängig sein:

$$(69) \quad E_{X,u}(X'u) = 0.$$

Unter den Annahmen S1 bis S7 behalten alle Ergebnisse (bis auf eine noch zu erwähnende Ausnahme) Gültigkeit, die wir im klassischen regressionsanalytischen Modell (A1 bis A3, K4, A5) dargestellt hatten, alle Schätzer bleiben erwartungstreu, die Bereichsschätzungen sind nach wie vor gültig. Dies ist der Fall, obwohl für die Verteilungsform der Prädiktoren keine Aussage gemacht wurde. So ist z.B.

$$(70) \quad \hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

nicht notwendig normalverteilt, wie dies im klassischen regressionsanalytischen Modell der Fall ist. Der Schätzer (70) ist übrigens in den Zufallsvariablen nichtlinear. Daraus resultiert die angekündigte Ausnahme: Er kann daher auch kein bester linearer erwartungstreuer Schätzer sein; er ist jedoch asymptotisch effizient (Schönfeld, 1971, p. 17).

Im korrelationsanalytischen Modell ist es auch natürlich, ein Maß für die Stärke des Zusammenhangs

$$(71) \quad \begin{aligned} \varrho^2 &= \varrho_{y(01\dots K)}^2 = \frac{\text{Spur } \Sigma_{yy} - N\sigma^2}{\text{Spur } \Sigma_{yy}} \\ &= \frac{\beta'[\mathbf{E}_x(\mathbf{X}'\mathbf{X}) - \frac{1}{N} \mathbf{E}_x(\mathbf{X}') \mathbf{J} \mathbf{E}_x(\mathbf{X})] \beta}{\beta'[\mathbf{E}_x(\mathbf{X}'\mathbf{X}) - \frac{1}{N} \mathbf{E}_x(\mathbf{X}') \mathbf{J} \mathbf{E}_x(\mathbf{X})] \beta + N\sigma^2} \end{aligned}$$

mit der Einsmatrix  $\mathbf{J}$ , d.h.  $(\mathbf{J})_{kl} = 1$ ,

in Form des multiplen Bestimmungsmaßes aufzustellen. Die Wurzel  $\varrho$  aus dem Bestimmungsmaß heißt *multiple Korrelation*. Da durch die Konstante  $x_0$  kein Zusammenhang vermittelt wird, lässt man sie häufig in der Notation der multiplen Korrelation weg:

$$(72) \quad \varrho = \varrho_{y(012\dots K)} = \varrho_{y(12\dots K)}.$$

Eine alternative Darstellungsform der multiplen Korrelation für den Spezialfall, daß Kriterium und (echte) Prädiktoren insgesamt  $(K+1)$ -dimensional normalverteilt sind, wird in Graybill (1976, p. 113, 418) beschrieben. Wählt man für die Prädiktoren eine Punktverteilung, was dem klassischen regressionsanalytischen Modell entspricht, ist die Definition (71) nach wie vor anwendbar, wobei die Symbole  $\mathbf{E}_x$  entfallen. In diesem Spezialfall wird besonders deutlich, daß der multiple Korrelationskoeffizient von den gewählten Prädiktorwerten abhängt; allgemein hängt er von den Charakteristika der Verteilung der Prädiktoren ab, wie natürlich auch von  $\beta$  und  $\sigma^2$ .

Als *Punktschätzung* für  $\varrho^2$  kann entsprechend Gleichung (71)

$$(73) \quad \begin{aligned} R^2 &= \frac{\mathbf{b}' \left[ \mathbf{X}'\mathbf{X} - \frac{1}{N} \mathbf{X}'\mathbf{J}\mathbf{X} \right] \mathbf{b}}{\mathbf{b}' \left[ \mathbf{X}'\mathbf{X} - \frac{1}{N} \mathbf{X}'\mathbf{J}\mathbf{X} \right] \mathbf{b} + (N-K-1)\delta^2} \\ &= \frac{\mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - \frac{1}{N} \mathbf{y}'\mathbf{J}\mathbf{y}}{\mathbf{y}'\mathbf{y} - \frac{1}{N} \mathbf{y}'\mathbf{J}\mathbf{y}} \end{aligned}$$

verwendet werden; dies ist genau gleich der schon in Gleichung (32) als empirische multiple Korrelation eingeführten Größe  $R^2$ . Mit der Matrix der empirischen Kovarianzen der Prädiktoren

$$(74) \quad S_{XX} = \frac{1}{N} (\mathbf{X}'\mathbf{X} - \frac{1}{N} \mathbf{X}'\mathbf{J}\mathbf{X}) = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & s_1^2 & s_{12} & \dots & s_{1K} \\ 0 & s_{21} & s_2^2 & \dots & s_{2K} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & s_{K1} & s_{K2} & \dots & s_K^2 \end{pmatrix},$$

dem Vektor der empirischen Kovarianzen von Prädiktoren und Kriterium

$$(75) \quad s_{XY} = \frac{1}{N} (\mathbf{X}'\mathbf{y} - \frac{1}{N} \mathbf{X}'\mathbf{J}\mathbf{y}) = \begin{pmatrix} 0 \\ s_{1y} \\ s_{2y} \\ \vdots \\ s_{Ky} \end{pmatrix}$$

und der Varianz des Kriteriums

$$(76) \quad s_{YY} = \frac{1}{N} (\mathbf{y}'\mathbf{y} - \frac{1}{N} \mathbf{y}'\mathbf{J}\mathbf{y}) = \bar{y}^2 - \bar{y}^2 = s_y^2$$

erhält man

$$(77) \quad R^2 = \frac{\mathbf{b}' S_{XX} \mathbf{b}}{s_{yy}} = \frac{s'_{XY} \mathbf{b}}{s_{yy}} = \frac{s'_{XY} S_{XX}^{-1} s_{XY}}{s_{yy}}.$$

Wenn die Konstante  $x_0$  aus den Matrizen und Vektoren ausgeschlossen wird, ändert sich am Ergebnis nichts, d.h. daß in (74) die erste Zeile und erste Spalte, in (75) die erste Komponente entfallen kann. Wenn für alle Variablen  $y$  und  $x_k$  eine multivariate Normalverteilung vorliegt, ist  $R^2$  der Maximum-Likelihood-Schätzer.  $R^2$  überschätzt im Mittel die Modell-Korrelation und ist demgemäß nicht erwartungstreu (siehe Olkin & Pratt, 1958). Wenn man die Schrumpfungskorrektur

$$(78) \quad \tilde{R}^2 = 1 - (1 - R^2) \frac{N-1}{N-K-1}$$

anwendet, ist der Bias dieser modifizierten Schätzung geringer. Für den Fall der einfachen Korrelation ist

$$(79) \quad \tilde{r} = r + \frac{r(1-r^2)}{2(N-4)}$$

eine gute Näherung für den unter der Voraussetzung der Binomialverteilung konstruierbaren (Olkin & Pratt, 1958; Graybill, 1976, p. 396) besten unverzerrten Schätzer von  $\rho$ .

Eine *Bereichsschätzung* für die multiple Korrelation  $\varrho$  im Fall der Multinormalverteilung ist in Graybill (1976, p. 419, 667) in tabellarischer Form angegeben. Für den Fall der einfachen Korrelation ( $K = 1$ ) nützt man die Tatsache aus, daß für die bivariate Normalverteilung

$$(80) \quad z = \frac{\operatorname{arctanh}(r) - \operatorname{arctanh}(\varrho)}{\sqrt{N - 3}}$$

näherungsweise standardnormalverteilt ist, um Bereichsschätzungen zu bilden. Ein näherungsweiser Konfidenzbereich zum Konfidenzgrad  $1 - \alpha$  ist daher durch

$$(81) \quad \tanh\left(\operatorname{arctanh}\left(r - \frac{N_{\alpha/2}}{\sqrt{N - 3}}\right)\right) \leq \varrho \leq \tanh\left(\operatorname{arctanh}\left(r + \frac{N_{\alpha/2}}{\sqrt{N - 3}}\right)\right)$$

gegeben.

Aus den genannten Bereichsschätzungen lassen sich wiederum *Signifikanztests* konstruieren. Die Nullhypothese  $\varrho^{(0)} = 0$  ist der Nullhypothese  $\beta_1^{(0)} = \beta_2^{(0)} = \dots = \beta_K^{(0)} = 0$  gleichwertig. Deswegen eignet sich zur Prüfung dieser Nullhypothese ebenso die Entscheidungsregel (34), die auch im korrelationsanalytischen Modell anwendbar ist.

Da viele Begriffsbildungen und Ergebnisse für das klassische regressionsanalytische Modell und das klassische korrelationsanalytische Modell gleichermaßen gelten, soll im folgenden vom *klassischen Modell* oder vom *Regressionsmodell* (manche Autoren sprechen vom Allgemeinen linearen Modell) die Rede sein, wenn es sich um gemeinsame Eigenschaften der beiden Ansätze handelt.

## 1.8 Modelle mit Fehlern in den Prädiktoren

Üblicherweise geht man in der Regressionsanalyse davon aus, daß die beobachtbaren Variablen exakt, d.h. ohne Fehler gemessen werden können. In der Regel sind die Variablen aber fehlerbehaftet; die Ursache liegt in zahlreichen Fehlerquellen, z.B. in der Erhebungstechnik (falsche Abgrenzungen, falsche Antworten), am Stichprobenfehler, an der Aufbereitung der Daten usw. Im folgenden werden Modelle mit fehlerbehafteten Variablen betrachtet. Als Literatur hierzu können Klein (1953, p. 282-305), Malinvaud (1966, p. 326-363), Cochran (1970), Schönfeld (1971, p. 100-124), Schneeweiß (1974, p. 216-232), Stroud (1974), Namboodiri et al. (1975), Wainer & Thyssen (1976), Aigner & Goldberger (1977), Maddala (1977, p. 292-319), Rock et al. (1977) und Dhrymes (1978, p. 242-268) herangezogen werden.

Es müssen nun für das Kriterium und jeden Prädiktor jeweils zwei Werte unterschieden werden: der richtige bzw. der tatsächlich beobachtete und damit

fehlerbehaftete. Um ersteren vom zweiten Wert in der Notation zu unterscheiden, wird er jeweils mit einem Stern versehen.

Aus schreibtechnischen Gründen wird nun für das Kriterium nicht mehr  $y$ , sondern  $s_0$  und für die Prädiktoren einschließlich der bisherigen Konstante  $x_0 = 1$  in abgeänderter Numerierung  $x_k$  geschrieben ( $1 \leq k \leq K$ ). Es bezeichnet also  $x_{0n}^*$  bzw.  $x_{0n}$  das tatsächliche bzw. beobachtete Kriterium zur Erhebungseinheit  $n$  und  $x_{kn}^*$  bzw.  $X_{kn}$  den tatsächlichen bzw. beobachteten  $k$ -ten Prädiktor zur Erhebungseinheit  $n$ .

Der Fehler in der Variablen  $X_{kn}$  ( $0 \leq k \leq K$ ,  $1 \leq n \leq N$ ) wird mit  $v_{kn}$  bezeichnet. Dieser Fehler kann additiv oder prozentual sein. Es wird ein additiver Zusammenhang unterstellt:

$$x_{kn} = x_{kn}^* + v_{kn} \quad (0 \leq k \leq K, 1 \leq n \leq N).$$

Die Regressionsbeziehung lautet:

$$x_{0n}^* = \sum_{k=1}^K \beta_k x_{kn}^* + \varepsilon_n \quad (1 \leq n \leq N),$$

wobei  $\varepsilon_n$  den Fehler in der Gleichung erfaßt.

In kompakter Matrizenform schreiben sich obige Beziehungen

$$(82) \quad X = X^* + V, \quad x_0 = x_0^* + v_0$$

$$(83) \quad x_0^* = X^* \beta + \varepsilon,$$

wobei  $X$  bzw.  $X^*$  bzw.  $V$  jeweils  $N \times K$ -Matrizen darstellen,  $\beta$  einen Spaltenvektor mit  $K$  Komponenten und  $x_0^*$  bzw.  $x_0$  bzw.  $\varepsilon$  Spaltenvektoren mit je  $N$  Komponenten. Wird  $x_0^*$  in (83) gemäß (82) ersetzt, so ergibt sich

$$(84) \quad x_0 = X^* \beta + \varepsilon + v_0.$$

Aus (84) resultiert, daß der Effekt des Fehlers in der Gleichung, nämlich  $\varepsilon$ , derselbe ist wie der des Fehlers in der Variable  $x_0$ , nämlich  $v_0$ . Wir setzen daher ohne Beschränkung der Allgemeinheit  $\varepsilon$  zu Null.

Wird weiter  $X^*$  in (84) gemäß (82) ersetzt, so ergibt sich die entsprechende Beziehung in den beobachtbaren Variablen

$$(85) \quad x_0 = X \beta + (v_0 - V \beta).$$

Der Fehler in der Gleichung (85) wird zusammengefaßt zu

$$u = v_0 - V \beta.$$

Damit resultiert die übliche Regressionsbeziehung

$$(86) \quad \mathbf{x}_0 = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}.$$

Folgende Annahmen die Fehler betreffend werden getroffen:

- (F1) Die Fehlervariable  $v_{kn}$  ist unabhängig von  $v_{lm}$  für  $n \neq m$  und  $0 \leq k, l \leq K$ .
- (F2)  $(v_{0n}, v_{1n}, \dots, v_{Kn})$  besitzt eine von  $n$  unabhängige Verteilung und  $\omega_{kl}$  bezeichnet die Kovarianz von  $v_{kn}$  und  $v_{ln}$  bzw.  $\Omega$  die entsprechende Kovarianzmatrix. Mit  $\Omega_{11}$  bezeichnen wir die Matrix, die aus  $\Omega$  durch Weglassen der ersten Zeile und Spalte entsteht, und mit  $\omega_1$  die erste Spalte von  $\Omega$  aber ohne  $\omega_{00}$ .
- (F3) Die Fehlervariablen  $v_{kn}$  sind alle unabhängig von den  $x_{ln}^*$  für  $0 \leq k, l \leq K, 1 \leq n \leq N$ .
- (F4)  $E v_{kn} = 0$  für  $0 \leq k \leq K, 1 \leq n \leq N$ .

Es wird zugelassen, daß einige der Variablen fehlerfrei beobachtbar sind, d.h. daß einige  $\omega_{kk}$  verschwinden können. Nur  $\omega_{00}$  sei ungleich Null, da in  $v_0$  auch der Fehler in der Gleichung miterfaßt wird.

Nun müssen noch die tatsächlichen Prädiktoren spezifiziert werden. Es gibt zwei Möglichkeiten:

1. Die stochastische Spezifikation, nach der alle  $x_{kn}^*$  ( $1 \leq k \leq K, 1 \leq n \leq N$ ) stochastische Variable sind (und damit auch  $x_{0n}^*$ );
2. Die deterministische Spezifikation, nach der alle  $x_{kn}^*$  fest vorgegeben sind. In diesem Falle ist wegen (83) (man beachte  $\epsilon=0$ ) auch  $x_{0n}^*$  fest vorgegeben.

Bei stochastischer Spezifikation soll noch gelten

$$(F5) \quad \lim_N \frac{1}{N} E(\mathbf{X}^{*'} \mathbf{X}^*) = \text{plim}_N \frac{1}{N} \mathbf{X}^{*'} \mathbf{X}^*,$$

und die Grenzwertmatrix  $M$  sei nichtsingulär.

Bei deterministischer Spezifikation soll gelten

$$(F5') \quad \lim_N \frac{1}{N} \mathbf{X}^{*'} \mathbf{X}^*$$

existiert und die Grenzwertmatrix sei nichtsingulär.

Zuerst wird unabhängig von der Spezifikation der Minimum-Quadrat-Schätzer für (86) betrachtet. Nach Definition von  $u$  und gemäß (86) ergibt sich

$$\mathbf{E} \mathbf{X}' \mathbf{u} = \mathbf{E} (\mathbf{X}^* - \mathbf{V})' (\mathbf{v}_0 - \mathbf{V} \boldsymbol{\beta}) = -\mathbf{E} \mathbf{V}' \mathbf{v}_0 + (\mathbf{E} \mathbf{V}' \mathbf{v}) \boldsymbol{\beta}.$$

Daraus ergibt sich (auch für den Fall, daß  $V$  und  $v_0$  unkorreliert sind) eine Korrelation der beobachtbaren Prädiktoren mit der Fehlervariablen  $u$ . Daher ist der Minimum-Quadrat-Schätzer

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{x}^*$$

verzerrt.

Gemäß (83) gilt aber (beachte  $\epsilon=0$ )

$$\mathbf{X}^{*'}\mathbf{X}^*\beta = \mathbf{X}^{*'}\mathbf{x}_0^*.$$

Darin ersetzen wir  $X^*$  und  $x_0^*$  gemäß (82). Es ergibt sich also

$$(\mathbf{X}-\mathbf{V})'(\mathbf{X}-\mathbf{V})\beta = (\mathbf{X}-\mathbf{V})'(\mathbf{x}_0-\mathbf{v}_0)$$

bzw. durch Ausmultiplizieren,

$$(87) \quad (\mathbf{X}'\mathbf{X}-\mathbf{V}'\mathbf{X}-\mathbf{X}'\mathbf{V}+\mathbf{V}'\mathbf{V})\beta = \mathbf{X}'\mathbf{x}_0-\mathbf{V}'\mathbf{x}_0-\mathbf{X}'\mathbf{v}_0+\mathbf{V}'\mathbf{v}_0.$$

Wegen der getroffenen Annahmen ergibt sich asymptotisch

$$\text{plim } \frac{1}{N}\mathbf{V}'\mathbf{X} = \text{plim } \frac{1}{N}\mathbf{X}'\mathbf{V} = \mathbf{0}$$

$$\text{plim } \frac{1}{N}\mathbf{V}'\mathbf{x}_0 = \text{plim } \frac{1}{N}\mathbf{X}'\mathbf{v}_0 = \mathbf{0}$$

$$\text{plim } \frac{1}{N}\mathbf{V}'\mathbf{V} = \boldsymbol{\Omega}_{11}$$

und

$$\text{plim } \frac{1}{N}\mathbf{V}'\mathbf{v}_0 = \boldsymbol{\omega}_1.$$

Daher ist das folgende Gleichungssystem asymptotisch äquivalent zu (87)

$$(88) \quad \left( \frac{1}{N}\mathbf{X}'\mathbf{X} + \boldsymbol{\Omega}_{11} \right) \tilde{\beta} = \frac{1}{N}\mathbf{X}'\mathbf{x}_0 + \boldsymbol{\omega}_1$$

Wenn also  $\boldsymbol{\Omega}_{11}$  und  $\boldsymbol{\omega}_1$  bekannt sind, so ergibt sich gemäß (88) aus den beobachtbaren Variablen der Schätzer  $\tilde{\beta}$ , der nach obiger Herleitung konsistent ist.

Dieser Schätzer setzt die Kenntnis von  $\boldsymbol{\Omega}_{11}$  und  $\boldsymbol{\omega}_1$  voraus. In vielen Fällen kann  $\boldsymbol{\omega}_1$  gleich Null gesetzt werden und  $\boldsymbol{\Omega}_{11}$  als diagonal angenommen werden, d.h. die Fehler in den verschiedenen Variablen sind untereinander unkorreliert. Trotzdem ist über die Beobachtungen hinaus eine gewisse a priori Kenntnis über die Varianzen der Fehler in den Variablen nötig, um die Parameter konsistent zu schätzen. Hier liegt ein Identifikationsproblem vor. Dieses wird

im Rahmen der stochastischen Spezifikation und bei Vorliegen von Normalverteilungen noch genauer diskutiert.

Die Struktur der Verzerrung des Minimum-Quadrat-Schätzers kann für den Fall der Einfachregression  $x_{0n}^* = \beta_1 + \beta_2 x_{2n}^*$  genauer studiert werden. Das Absolutglied  $x_{1n} = 1$  kann natürlich ohne Fehler beobachtet werden. Weiter sei  $\omega_{01} = 0$ .

Damit ergibt sich

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{x}_0$$

bzw.

$$\text{plim } \hat{\beta} = \left( \text{plim } \frac{1}{N} \mathbf{X}^{*'} \mathbf{X}^* + \begin{pmatrix} 0 & 0 \\ 0 & \omega_{22} \end{pmatrix} \right)^{-1} \left( \text{plim } \frac{1}{N} \mathbf{X}^{*'} \mathbf{X}^* \boldsymbol{\beta} \right)$$

und

$$\text{plim } \hat{\beta}_2 = \beta_2 \frac{1}{1 + \frac{\omega_{22}}{s}}$$

$$\text{wobei } s = \text{plim } \frac{1}{N} \sum_n (x_{2n}^* - \bar{x}_2^*)^2, \bar{x}_2^* = \frac{1}{N} \sum_n x_{2n}^* .$$

Offensichtlich wird  $\beta_2$  immer unterschätzt und zwar um so mehr, je größer das Verhältnis der Varianz von  $v_{2n}$  zu  $x_{2n}^*$  ist. Wenn wir z.B.  $N = 10$  annehmen und  $x_{2n}^* = c_1$  für  $1 \leq n \leq 5$  bzw.  $x_{2n}^* = c_2 \neq c_1$  für  $6 \leq n \leq 10$ , so ist es relativ klein. Grafisch ergibt sich die Darstellung in Abb. 2.

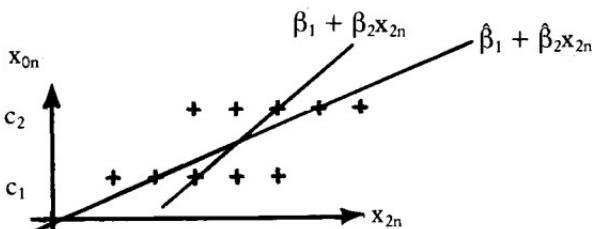


Abb. 2: Verzerrung der Regressionskoeffizienten, wenn der Prädiktor mit Meßfehler gemessen wurde.

Es seien nun alle Fehlervariablen  $v_{kn}$  und alle  $x_{kn}^*$  für  $0 \leq k \leq K$ ,  $1 \leq n \leq N$  normalverteilt. Mit  $\mu_k$  wird der Erwartungswert von  $x_{kn}^*$  bezeichnet ( $0 \leq k \leq K$ ). Es gilt dann wegen (83)

$$(89) \quad \mu_0 = \sum_{k=1}^K \mu_k \beta_k$$

Mit  $\Sigma = (\sigma_{kl})$  wird die  $(K+1) \times (K+1)$ -Kovarianz-Matrix von  $(x_{0n}, \dots, x_{kn})$  bezeichnet und mit  $\Sigma^* = (\sigma_{kl}^*)$  die  $K \times K$ -Kovarianz-Matrix von  $(x_{1n}^*, \dots, x_{kn}^*)$ .

Es gilt dann

$$\sigma_{kl} = \sigma_{kl}^* + \omega_{kl} \text{ für } 1 \leq k, l \leq K$$

$$\sigma_{00} = \sum_{k,l=1}^K \beta_k \sigma_{kl}^* \beta_l + \omega_{00}$$

$$\sigma_{0l} = \sum_{k=1}^K \beta_k \sigma_{kl}^* + \omega_{0l} \text{ für } 1 \leq l \leq K.$$

Mit diesen Bezeichnungen ergibt sich für die Log-Likelihood-Funktion (bis auf einen unwesentlichen konstanten Term)

$$-\frac{N}{2} \ln(\det \Sigma) - \frac{1}{2} \sum_{n=1}^N \sum_{k,l=0}^K (x_{kn} - \mu_k) \sigma_{kl}^{(-1)} (x_{ln} - \mu_l).$$

Aus den beobachtbaren Variablen können wir bestenfalls die Matrix  $\Sigma$  und  $\mu_0, \dots, \mu_K$  erhalten. Die eigentlich interessierenden Parameter sind aber  $\mu_1, \dots, \mu_K, \beta, \Sigma^*$  und  $\Omega$ .

Offensichtlich sind letztere viel zahlreicher und damit nicht alle identifizierbar. Durch geeignete a priori Information z.B., daß  $\Omega$  diagonal ist und gewisse Hauptdiagonalelemente Null sind, kann Identifizierbarkeit erreicht werden.

Für den Fall der deterministischen Spezifikation kann eine andere Schätzfunktion hergeleitet werden. Die ersten  $K_1 \leq K$  Variablen seien fehlerbehaftet und die restlichen exakt beobachtbar. Diese  $K_1$  ersten Variablen seien als  $X_1^*$  und  $V_1$  definiert. Damit gilt nun anstatt (82) und (83)

$$(82') \quad X_1 = X_1^* + V_1$$

und

$$(83') \quad x_0^* = X^* \beta$$

Sei  $\Omega_1$  die nichtsinguläre Kovarianz-Matrix der Zeilen von  $V_1$ , dann wird folgende gewichtete Minimum-Quadrat-Methode angewandt

$$\text{Spur } (X_1 - X_1^*) \Omega_1^{-1} (X_1 - X_1^*)^\top = \text{Minimum}$$

unter der Nebenbedingung (83'). Dabei werden die Größen  $x_{kn}^*$  als Parameter angesehen, die „mitgeschätzt“ werden. Der resultierende Schätzer ist konsistent.

## 1.9 Zeitreihenanalyse im allgemeinen regressionsanalytischen Modell

In neuerer Zeit gewinnt in der Psychologie die auch quantitative Untersuchung von zeitlichen Prozessen größere Bedeutung, mit der die Untersuchung von Zuständen ergänzt werden kann. Dies hat wohl nicht zuletzt seinen Grund darin, daß wichtige Teile des quantitativ-methodischen Instrumentariums für die speziellen Bedürfnisse der Veränderungsbeschreibung erst jüngst entwickelt (Yule, 1927; Hannan, 1955; Hannan, 1960; Box und Tiao, 1965; Kendall & Stuart, 1966, p. 403-430; Malinvaud, 1966, p. 373-393, 425-453; Box et al., 1967; Box & Jenkins, 1970; Box & Pierce, 1970, Box & Tiao, 1975; Glass et al., 1975; Anderson, 1976; Kendall, 1976) oder in die psychologische Literatur (Holtzmann, 1963; Huber, 1967; Huber, 1973; Dahme, 1975; Dahme, 1976; Gudat & Revenstorff, 1976; Hersen & Barlow, 1976; Petermann, 1977a, 1977b, 1978; Revenstorff, 1979; Revenstorff & Keeser, 1979; McCleary & Hay, 1980; Keeser, 1981) aufgenommen wurden.

Das klassische regressionsanalytische Modell

$$y = X\beta + u$$

mit K4, wozu die Unkorreliertheit der Komponenten  $u_n$  und  $u_m$  für  $n \neq m$ ,  $1 \leq n, m \leq N$ , gehört, ist für Zeitreihenanalysen deswegen nicht geeignet, weil in der Regel der nicht durch  $X$  erklärbare Anteil des Kriteriums, die Fehlervariable  $u$ , durch nicht erfaßte, zeitlich träge Größen beeinflußt wird. Bei einem Experiment zur Rauchertherapie wird zwar einerseits das Kriterium  $y$ , die Anzahl gerauchter Zigaretten, wesentlich durch die in den Prädiktorvariablen  $x_k$  dargestellten Interventionen bestimmt sein, andererseits werden sich zeitlich kurzfristig und längerfristig wirksame Einflußgrößen bemerkbar machen, deren Wirkung auf das Kriterium durch die Fehlervariable  $u$  zusammengefaßt wird. Beispielsweise wird ein zum Zeitpunkt  $t_n$  auftretender Partnerkonflikt eines Klienten sich so auswirken, daß für die Anzahl gerauchter Zigaretten die Fehlerwerte  $u_n, u_{n+1}, \dots$  mit sich abschwächender Tendenz systematisch erhöht sind. Insofern wird das Kriterium nun von drei Varianzquellen bestimmt, den Prädiktoren, den Fehlern und der seriellen Abhängigkeit der Fehler. Unter bestimmten Umständen, die an einem Beispiel noch näher erläutert werden, können derartige Zusammenhänge durch den Ansatz einer von der Einheitsmatrix I verschiedenen Matrix V im allgemeinen regressionsanalytischen Modell berücksichtigt werden.

Leider ist es so, daß in praktischen Fällen nicht einmal die Matrix V als Indikator der Struktur der Kovarianzmatrix der Zufallsvariablen  $u_n$

$$(90) \quad \Sigma_{uu} = \sigma^2 V = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2N} \\ \vdots & \vdots & & \\ \sigma_{N1} & \sigma_{N2} & \dots & \sigma_{NN} \end{pmatrix}$$

bekannt ist. Die Matrix  $\Sigma_{uu}$  muß daher geschätzt werden, wobei es ausreicht, die wechselseitigen Verhältnisse der Kovarianzen zu schätzen, wie sie in  $V$  erfaßt sind. Leider hat man zum Schätzen der  $(N(N+1)/2)-1$  freien Parameter in  $V$  nur die  $N$  Werte  $y_n$  zur Verfügung, so daß eine konsistente Schätzung der Matrix  $V$  so nicht möglich ist. Die Zahl der freien Parameter in  $\Sigma_{uu}$  muß also dadurch verringert werden, daß ein angemessenes stochastisches Modell mit erheblich weniger freien Parametern angesetzt wird, das Beziehungen zwischen den Kovarianzen  $\sigma_{nm}$  definiert.

Ist  $V$  geschätzt, kann in einer zweiten Phase die Schätzung der Regressionskoeffizienten und  $\sigma^2$  nach den in Kap. 1.3 beschriebenen Methoden erfolgen. Leider sind die dort angegebenen Eigenschaften der Schätzer nicht für die hier beschriebene *zweiphasige* Schätzung (two step estimator) gültig. Unter welchen Bedingungen andere, nur asymptotische Schätzegenschaften nachgewiesen werden können, beschreibt etwa Schönfeld (1969, p. 207). Auch bei Vorliegen der erforderlichen Normalverteilungsannahme sind die in Kap. 1.3 berichteten Bereichsschätzungen und die daraus abgeleiteten Signifikanztests nur asymptotisch, für endliche Zeitreihen also nur näherungsweise gültig. Natürlich ist dabei die Näherung erheblich besser, als wenn die Abhängigkeit der Fehler ignoriert und fälschlich das klassische regressionsanalytische Modell angenommen wird (Keeser, 1981, p. 281). Wie sehr das Signifikanzniveau

Tabelle 1: Verzerrung der Irrtumswahrscheinlichkeit bei autokorrelierten Daten; relative Häufigkeiten von signifikanten Ergebnissen auf dem 5 %-Niveau von jeweils 1000 Stichproben von Zufallszahlen beim t-Test und Mann-Whitney-Test für verschiedene Verteilungsformen und verschiedene Werte der Autokorrelation (nach Box, 1974).

Q	Test	Verteilungsform		
		Normal	Rechteck	$\chi^2$ (schief)
0.0	t-Test	.054	.056	.047
0.0	M-W-Test	.045	.043	.043
- .4	t-Test	.003	.005	.001
- .4	M-W-Test	.001	.005	.002
+ .4	t-Test	.105	.125	.114
+ .4	M-W-Test	.096	.110	.101

einfacher t-Tests (der t-Test ist der Test (37)) oder von Mann-Whitney-Tests verfälscht wird, wenn die Fehler autoregressiv korreliert sind (siehe unten), zeigen die Ergebnisse einer Simulationsstudie von Box (1974), die in Tab. 1 wiedergegeben ist.

Für die Anwendung des zweiphasigen Schätzers ist ein nichttriviales Problem das Aufstellen eines angemessenen und zugleich sparsamen stochastischen Modells für die Fehlervariablen, das in der Praxis anhand der vorliegenden Daten erfolgen muß. Eine Klasse von möglichen Modellen wird unter der Bezeichnung ARIMA-Modell (autoregressive-integrated moving average model) zusammengefaßt. Die Eigenschaften des Modells, die Methoden zur Identifikation des speziell für die Daten angemessenen Modells und die Schätzung der Parameter des Modells stellen z.B. McCleary & Hay (1980) und Keeser (1981) ausführlich dar. Computerprogramme zur Bearbeitung dieser Aufgabe sind in den neueren Versionen der Programme BMDP2T (Dixon & Brown, 1981), SPSS-Prozedur BOX-JENKINS (Beutel & Schubö, 1982) und in der Unterprogrammbibliothek IMSL (IMSL, 1979) enthalten.

Als einfachstes und am besten dokumentiertes (z.B. Johnston, 1963, p. 177; Schönfeld, 1969, p. 211; Schönfeld, 1971, p. 31; Schneeweiß, 1974, p. 180; Seifert, 1975, p. 109) Beispiel sollen die Ergebnisse beim autoregressiven Schema *erster Ordnung* (AR(1) oder ARIMA (1,0,0)) vorgestellt werden, woran das Prinzip der Vorgehensweise sichtbar wird. Dies ist der einfachste Ansatz, der jedoch leider für die meisten Anwendungen z.B. in der Psychotherapieforschung nicht ausreicht.

Als stochastisches Modell für die Fehlervariable wird angenommen:

- (Z1)  $u_n = \varrho u_{n-1} + e_n$  für  $n = 2, 3, \dots, N$  mit
- (Z2)  $|\varrho| < 1$ ,  $E u_1 = 0$ ,  
wobei für die Zufallsvariablen  $e_n$  gilt
- (Z3)  $E e_n = 0$  für  $n = 1, 2, \dots, N$
- (Z4)  $\Sigma_{ee} = \sigma_e^2 I$
- (Z5)  $\text{cov}(u_{n-1}, e_n) = 0$  für  $n = 2, 3, \dots, N$

Es besteht also ein linearer Zusammenhang zwischen den Fehlervariablen zu unmittelbar aufeinanderfolgenden Zeitpunkten  $t_{n-1}$  und  $t_n$ , der selbst wieder von unregelmäßigen Zufallsschocks überlagert wird. Für die Beziehung aufeinanderfolgender Fehlervariablen wird demnach ein klassisches korrelationsanalytisches Modell angesetzt. Unter den Bedingungen Z1 bis Z5 folgt  $y_n$  einem stationären Prozeß. Es gilt

$$(91) \quad u_n = \varrho^{n-1} u_1 + \varrho^{n-2} e_2 + \dots + \varrho e_{n-1} + e_n \quad \text{für } n = 2, 3, \dots, N$$

woraus sich wegen  $|\varrho| < 1$  ablesen läßt, daß die Zufallsschocks für spätere

Zeitpunkte einen immer schwächer werdenden Einfluß haben. Die Kovarianzmatrix der Fehlervariablen hat die Form

$$(92) \quad \Sigma_{uu} = \sigma^2 V \text{ mit} \\ \sigma^2 = \sigma_u^2 = \frac{\sigma_e^2}{1-\rho^2} \text{ und}$$

$$V = \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{N-1} \\ \rho & 1 & \rho & \dots & \rho^{N-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho^{N-1} & \rho^{N-2} & \rho^{N-3} & \dots & 1 \end{pmatrix}$$

Die für die Schätzung der Regressionskoeffizienten benötigte Inverse ist

$$(93) \quad V^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho & 0 & \dots & 0 & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \dots & 0 & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\rho & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \dots & 0 & -\rho & 1 \end{pmatrix}$$

Kennt man die Autokorrelation  $\rho$ , dann lassen sich damit in der zweiten Phase innerhalb des allgemeinen regressionsanalytischen Modells alle Schätzungen, Prognosen und Tests exakt durchführen, vorausgesetzt die Normalverteilungsannahme A5 ist erfüllt.

Im Falle des autoregressiven Schemas erster Ordnung (Wonnacott & Wonnacott, 1970, p. 328) wie im allgemeinen ARIMA-Modell (Keeser, 1981, p. 207) läßt sich die sog. Aitken-Reduktion auf das klassische regressionsanalytische Modell auf besonders einfache Weise durch Differenzbildung durchführen. Man zerlegt  $V^{-1}$  in zwei Faktoren

$$(94) \quad V^{-1} = D'D \text{ mit}$$

$$D = \frac{1}{\sqrt{1-\rho^2}} \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \dots & 0 & 0 \\ -\rho & 1 & 0 & \dots & 0 & 0 \\ 0 & -\rho & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & -\rho & 1 \end{pmatrix}$$

und erhält in

$$(95) \quad y^* = X^* \beta + u^*$$

mit  $y^* = Dy$ ,  $X^* = DX$ ,  $u^* = Du$

ein klassisches regressionsanalytisches Modell mit unkorrelierten Fehlervariablen  $(\mathbf{D}\mathbf{u})_n$ . Daß es sich bei  $\mathbf{D}$  um eine Matrix handelt, die Differenzen aufeinanderfolgender Messungen bildet, sei an

$$(96) \quad \mathbf{Dy} = \begin{bmatrix} \sqrt{1-\varrho^2} & y_1 \\ y_2 - \varrho y_1 & \\ y_3 - \varrho y_2 & \\ \vdots & \\ y_N - \varrho y_{N-1} & \end{bmatrix}$$

demonstriert: wie man sieht, muß die erste Messung  $y_1$  etwas abweichend behandelt werden. Es genügt für die Transformation in das klassische Modell übrigens nicht, nur die Kriteriumswerte zu transformieren, auch die Prädiktorwerte müssen, wie aus (95) ersichtlich, mittransformiert werden.

Bisher wurde davon ausgegangen, daß die Autokorrelation  $\varrho$  bekannt ist. Im allgemeinen muß man sie jedoch schätzen (Phase 1). Dazu bildet man die OLS-Schätzungen (klassisches Regressionsmodell) für die Residuen

$$(97) \quad \hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

und schätzt  $\varrho$  durch

$$(98) \quad \hat{\varrho} = \frac{\sum_{n=2}^N \hat{u}_{n-1} \hat{u}_n}{\sum_{n=1}^{N-1} \hat{u}_n}$$

Dieser Schätzer ist nicht erwartungstreu. Die Schätzung setzt man anstelle des Parameters  $\varrho$  in den Gleichungen (92) bis (96) für die zweite Schätzphase ein, die sonst wie in dem Fall erfolgt, in dem  $\varrho$  a priori bekannt ist.

Bestehen Zweifel, ob überhaupt Autokorrelation der Residuen vorliegt, kann die Entscheidung darüber auch durch statistische Tests herbeigeführt werden. Pauschal kann etwa mit einem Runs-Test auf die geschätzten Residuen (97) entschieden werden, ob die Vorzeichen der Residuen tatsächlich regellos verteilt sind. Für die spezifische Prüfung der Nullhypothese, daß die Autokorrelation  $\varrho = 0$  ist, müssen zwei Varianten unterschieden werden, die beide die Normalverteilungsannahme A5 erfordern.

Für die erste Variante müssen die beiden Residuen selbst bekannt sein, d.h. die Regressionskoeffizienten  $\beta$  müssen bekannt sein. Dann ist die Verteilung des von Neumann-Verhältnisses

$$(99) \quad Q(y) = \frac{\frac{1}{N-1} \sum_{n=2}^N (u_n - u_{n-1})^2}{\frac{1}{N} \sum_{n=1}^N u_n^2}$$

unter der Nullhypothese bekannt (Tabellen in Hart & von Neumann, 1942).  $Q$  ist symmetrisch um  $2N/(N - 1)$  mit Varianz  $\text{var } Q \approx 4/N$  verteilt.

Für den realistischeren Fall, daß nur die Schätzungen der Residuen nach (97) vorliegen, gibt es die zweite Variante, den *Durbin-Watson-Test* (Durbin & Watson, 1950, 1951, 1971), der auf der Prüfgröße

$$(100) \quad d = \frac{\sum_{n=2}^N (\hat{u}_n - \hat{u}_{n-1})^2}{\sum_{n=1}^N \hat{u}_n^2}$$

beruht. Es handelt sich dabei um einen „bounds-Test“, bei dem untere und obere Schranken für die Durbin-Watson-Statistik tabellarisch erfaßt sind (z.B. Koerts & Abrahamse, 1969). Die Tabellen sind in der Regel für die einseitige Alternativhypothese  $\rho > 0$  ausgelegt. Unterschreitet  $d$  die untere Schranke  $d_{L,\alpha}$ , wird die Nullhypothese zugunsten der Alternativhypothese abgelehnt. Wird die obere Schranke  $d_U$ , überschritten, entscheidet man sich für die Nullhypothese. Liegt  $d$  zwischen  $d_{L,\alpha}$  und  $d_{U,\alpha}$ , erhält man sich des Urteils (s. Abb. 3).

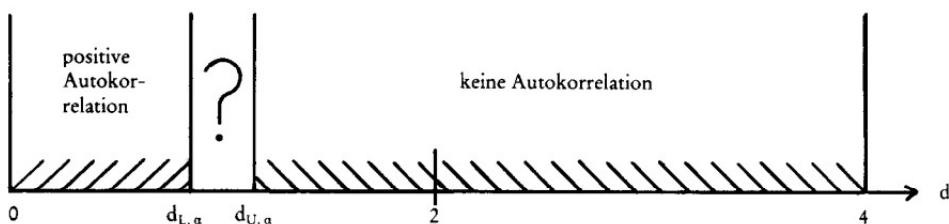


Abb. 3: Durbin-Watson-Test auf positive Autokorrelation.

Die beiden besprochenen Tests sind für die Diagnostik des autoregressiven Prozesses erster Ordnung konstruiert, der in Anwendungsfällen der Psychologie eher selten vorzufinden ist. Für allgemeinere serielle Abhängigkeiten haben Box & Pierce (1970) einen Modellanpassungstest entwickelt, den sie Portemanteau-Test nennen. Zur Entscheidung, ob die Residuen einem periodischen Muster folgen, empfehlen Jenkins & Watts (1968) den Periodogramm-Test.

Wenn auch die Prädiktoren selbst Zufallsvariablen mit seriellen Abhängigkeiten sind, müssen Transfermodelle (distributes lags, dynamische Regression) herangezogen werden (s. Box & Jenkins, 1970; McCleary & Hay, 1980).

## 1.10 Suppression und Kollinearität

Die Interpretation von Regressionsanalysen ist deswegen häufig nicht einfach, weil die Regressionskoeffizienten stark von den Interkorrelationen der Prädiktoren abhängen. Die Schwierigkeiten bei der Interpretation einer Regressionsanalyse werden an einem Beispiel aufgezeigt, dessen Bestimmungsgrößen wir später variieren. Dabei werden die Begriffe Wichtigkeit einzelner Prädiktoren, Suppression und Kollinearität eingeführt und diskutiert.

Cohen & Cohen (1975, p. 73) stützen sich bei der Diskussion der Zerlegung der Varianz des Kriteriums, die durch eine Regressionsanalyse gegeben ist, auf ein Beispiel, das wir zur Illustration übernehmen wollen. Es wird eine Stichprobe von Mitgliedern des Lehrkörpers von Universitäten untersucht. Als Kriteriumsvariable  $y$  dient der akademische Rang einer Person, der, übertragen auf deutsche Verhältnisse, dadurch skaliert wird, daß von der Besoldungsgruppe H1 bis H4 der Buchstabe entfernt wird; Mitgliedern des akademischen Mittelbaus könnte wahlweise der Wert 1 oder 0 zugewiesen werden. Der Zusammenhang zwischen diesem Kriterium und den beiden Prädiktoren  $x_1 =$  Anzahl der Publikationen und  $x_2 =$  akademisches Alter = Anzahl der Jahre seit der Promotion soll untersucht werden. Wenn die Personen der Stichprobe zufällig aus der Grundgesamtheit entnommen werden, liegt ein klassisches korrelationsanalytisches Modell vor; wegen der Zufallsreihenfolge in der Stichprobe sind serielle Korrelationen ausgeschlossen. Falls für die Argumentation erforderlich, wollen wir für die Residuen Normalverteilung annehmen,

Tabelle 2: Die Ergebnisse zum Beispiel von Cohen & Cohen (1975, p. 74). Der Stichprobenumfang ist  $N = 15$ .

Variable	Akad. Rang $y$	Konstante $x_0$	Anz. Publik. $x_1$	Akad. Alter $x_2$
Mittelwert	1.8	1.0	7.6	9.6
Standardabw. s	.748	.0	4.96	7.25
einf. Korr. mit $y$	1.0	—	.463	.612
einf. Korr. mit $x_1$	.463	—	1.0	.683
Regressionskoeff. $b_k$	—	1.15	.013	.057
Regressionskoeff. $b_k^*$	—	.0	.084	.555
für stand. Variablen				

obwohl sie wegen der Diskretheit des Kriteriums und der Beschränkung des Wertebereichs sicher höchstens in schlechter Näherung gegeben sein kann. Die meisten Interpretationen benötigen jedoch die Normalverteilungsannahme nicht. Außerdem sollen zwischen allen drei Variablen nur lineare Zusammenhänge bestehen. Die wichtigsten Ergebnisse der empirischen Regression mit den Schätzungen der Regressionskoeffizienten im klassischen Modell sind in Tab. 2 enthalten.

In dieser Tabelle sind neben den Regressionskoeffizienten aus

$$(101) \quad \hat{y}_n = b_0 + b_1 x_{n1} + b_2 x_{n2} \quad (1 \leq n \leq N)$$

auch die Regressionskoeffizienten für die standardisierten Werte

$$(102) \quad z_{yn} = \frac{y_n - \bar{y}}{s_y}, \quad z_{nk} = \frac{x_{nk} - \bar{x}_k}{s_k} \quad (1 \leq n \leq N, 1 \leq k \leq K)$$

enthalten. Die Regressionskoeffizienten  $b_k^*$  für standardisierte Werte (in der Literatur werden sie häufig auch Beta-Gewichte genannt) sind deshalb angenehmer für die Interpretation, weil sie nicht von der Skalierung, d.h. von den Maßeinheiten der Variablen abhängen. Sie sind die Bestimmungsgrößen zur OLS-Schätzung der standardisierten Kriteriumswerte

$$(103) \quad \hat{z}_{yn} = b_1^* z_{n1} + b_2^* z_{n2} \quad (1 \leq n \leq N),$$

wobei die Variable  $x_0$  nicht mehr berücksichtigt werden muß.

In Abb. 4 werden die Anteile gemeinsamer Varianz durch das Überlappen von Kreisscheiben dargestellt und mit den entsprechenden einfachen semipartialen (Cohen & Cohen, 1975, p. 81; Gaenslen & Schubö, 1976, p. 88) oder multiplen Korrelationen etikettiert. Die quadrierten Korrelationskoeffizienten sind zugleich die erklärten Varianzen, wenn es sich um standardisierte Variable handelt.

Richtig wiedergegeben wird in Abb. 4, daß für Fläche a verschiedene äquivalente Berechnungsmöglichkeiten bestehen

$$(104) \quad \begin{aligned} a &= R_{y(12)}^2 - r_{y(1-2)}^2 - r_{y(2-1)}^2 \\ &= r_{y1}^2 - r_{y(1-2)}^2 \\ &= r_{y2}^2 - r_{y(2-1)}^2 \\ &= .2106. \end{aligned}$$

Doch ist die Abbildung insofern gefährlich, als sie nicht berücksichtigt, daß dieser Betrag bei anderen Daten durchaus negativ sein könnte (s. Tab. 3). Auch können in einer derart vereinfachten Abbildung häufig die Beziehungen der Prädiktoren untereinander nicht angemessen dargestellt werden.

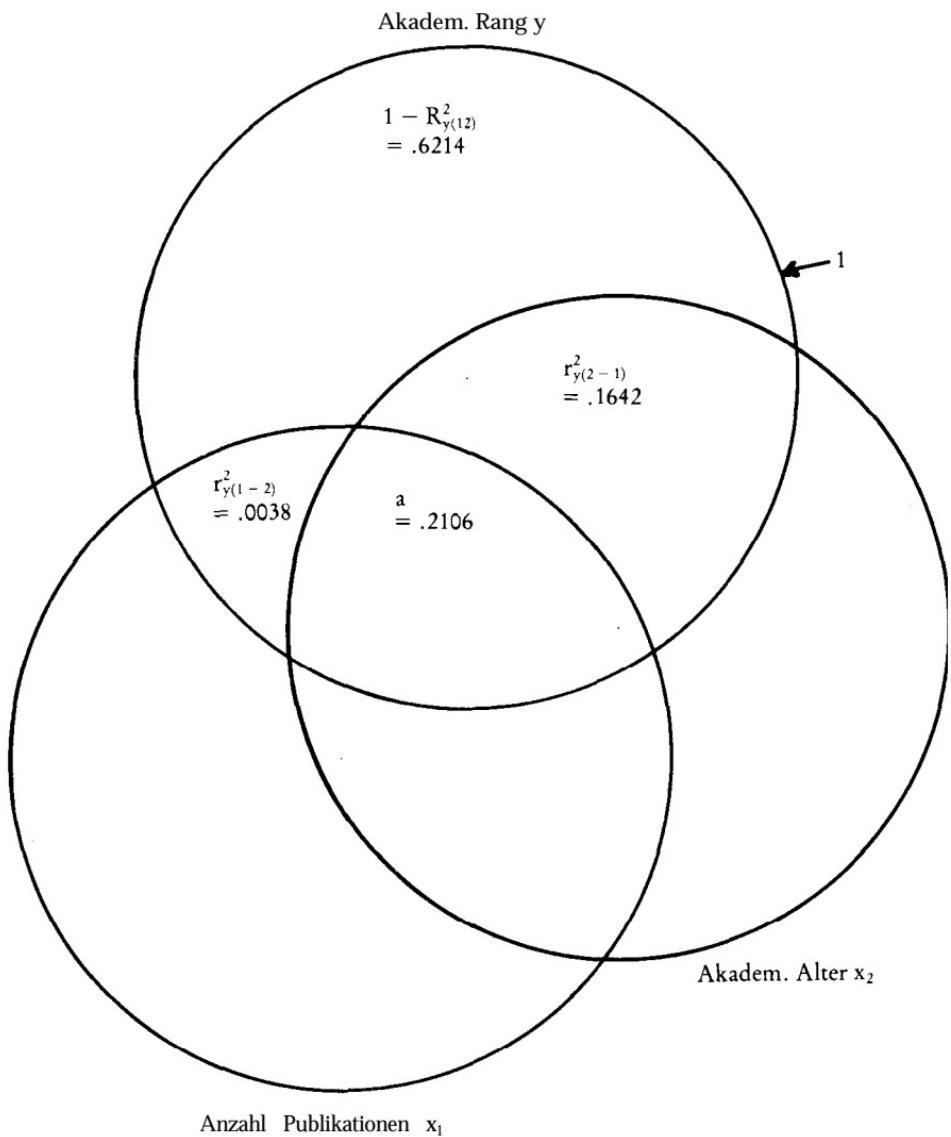


Abb. 4: Die Varianzzerlegung bei der Regression mit zwei Prädiktoren im Beispiel von Cohen & Cohen (1975, p. 74).

Wie lässt sich dieses Ergebnis nun deskriptiv interpretieren? Einerseits besteht ein Zusammenhang zwischen der Anzahl der Publikationen und dem akademischen Rang, der wechselseitig 21% der Varianz erklärt; über eine kausale gerichtete Abhängigkeit lässt sich bei Querschnittsuntersuchungen nie etwas

aus den Daten, sondern allenfalls aufgrund einer Theorie über den Gegenstandsbereich aussagen. Andererseits besteht ein noch stärkerer Zusammenhang (erklärter Varianzanteil 37%) zwischen dem akademischen Alter und dem akademischen Rang. Die Betrachtung des gemeinsamen Zusammenhangs der Variablen Anzahl der Publikationen und akademisches Alter mit dem akademischen Rang wirft gegenüber der isolierten Betrachtung der Prädiktoren ein neues Licht vor allem auf die Rolle der Anzahl der Publikationen. Der durch sie zusätzlich erklärte Anteil der Kriteriumsvarianz beträgt nämlich nur  $r_{y(1-2)}^2 = .0038$  und ist demnach verschwindend gering. In der Begriffsbildung von Bortz (1977, p. 596) ist die Anzahl der Publikationen demnach eine redundante Prädiktorvariable. Das bedeutet, daß für die Schätzung des akademischen Rangs die Hinzunahme der Anzahl der Publikationen zum akademischen Rang keine Verbesserung erbringt. Dies zeigt sich übrigens nicht so deutlich an den Regressionskoeffizienten  $b_k^*$ , wo  $x_1$  mit 0.084 gewichtet ist. Durch den Zusammenhang zwischen  $x_1$  und  $x_2$  entsteht eine schwache Kollinearität, die dazu führt, daß aus beiden Prädiktoren gleichermaßen schätzbare Anteile des Kriteriums auf die Prädiktoren aufgeteilt geschätzt werden.

Um zu demonstrieren, wie sehr die Ergebnisse von der Korrelation der Prädiktoren unter sonst gleichen Umständen abhängen, werden einige für die Interpretation einer Regressionsanalyse wichtige Größen in Abhängigkeit von der Prädiktorenkorrelation dargestellt. Die Regressionskoeffizienten für standardisierte Variablen können im hier behandelten Fall zweier Prädiktoren nach

$$(105) \quad b_1^* = \frac{r_{y1} - r_{y2}r_{12}}{1 - r_{12}^2}, \quad b_2^* = \frac{r_{y2} - r_{y1}r_{12}}{1 - r_{12}^2},$$

die quadrierten semipartialen Korrelationen, die den zusätzlichen Varianzbeitrag der Variablen angeben, nach

$$(106) \quad r_{y(1-2)} = \frac{r_{y1} - r_{y2}r_{12}}{\sqrt{1 - r_{12}^2}}, \quad r_{y(2-1)} = \frac{r_{y2} - r_{y1}r_{12}}{\sqrt{1 - r_{12}^2}}$$

und der insgesamt erklärte Varianzanteil nach

$$(107) \quad R_{y(12)}^2 = \frac{(r_{y1} - r_{y2})^2}{1 - r_{12}^2} + 2 \cdot \frac{r_{y1}r_{y2}}{1 + r_{12}}$$

berechnet werden. Aus der Bedingung, daß  $R_{y(12)}^2$  den Wert 1 nicht übersteigen darf, ergibt sich der zulässige Wertebereich für  $r_{12}$

$$(108) \quad r_{y1}r_{y2} - \sqrt{(1 - r_{y1}^2)(1 - r_{y2}^2)} \leq r_{12} \leq r_{y1}r_{y2} + \sqrt{(1 - r_{y1}^2)(1 - r_{y2}^2)}.$$

Für die Grenzen dieses Intervalls wird der insgesamt erklärte Varianzanteil

stets 1. Der minimal erklärte Varianzanteil wird, wenn o.B.d.A.  $|r_{y1}| \leq |r_{y2}|$ ,

$$(109) \quad r_{12} = \frac{r_{y1}}{r_{y2}}$$

angenommen und sein Betrag ist dann

$$R^2_{y(12)} = r_{y2}^2 .$$

Die konkreten Ergebnisse für die im Beispiel vorliegende Prädiktorkorrelation und einige weitere Korrelationswerte sind in Tab. 3 wiedergegeben. Es überrascht, wie stark die Ergebnisse voneinander abweichen, obwohl die Stärke des Zusammenhangs jedes einzelnen Prädiktors zum Kriterium stets dieselbe ist.

Leider sind die Ergebnisse nur für zwei Fälle leicht zu interpretieren: Zum ersten im Fall der minimal erklärten Varianz,  $r_{12} = 0.463/0.612 = .757$ , bei dem die Schätzung des Kriteriums ausschließlich durch den damit am stärksten korrelierenden Prädiktor erfolgt, und zum zweiten im Fall  $r_{12}=0$ , wo die Prädiktoren unkorreliert sind. In allen anderen Fällen ist es prinzipiell unmöglich, befriedigende Maßzahlen für den alleinigen Beitrag einer Prädiktorvariablen innerhalb der gemeinsamen Schätzung herauszunehmen, so sehr der Praktiker sich Hinweise auf die Wichtigkeit oder Bedeutung einzelner Prädiktoren wünscht. Dennoch werden als Indizes der *Wichtigkeit einzelner Prädiktoren* die einfachen Korrelationen mit dem Kriterium, die Regressionskoeffizienten für standardisierte Variablen, die quadrierten semipartialen Korrelationen oder die Produkte von einfacher Korrelation und Regressionskoeffizient für standardisierte Variablen herangezogen. Alle diese Größen haben ihre sinnvolle Bedeutung, doch sind sie alle nur in den beiden genannten Fällen der leichten Interpretierbarkeit als Indizes für den eigenständigen Beitrag eines Prädiktors uneingeschränkt verwendbar. Die spezifischen Vor- und Nachteile dieser Größen werden am ausführlichsten von Darlington (1968) diskutiert.

Recht kontrovers wird in der Literatur auch die Frage diskutiert, was unter *Suppressorvariablen* verstanden werden soll, und wie sie noch näher charakterisiert und gegliedert werden können. Gemeinsam ist die Forderung, daß eine Suppressorvariable eine Variable sein soll, die im Kontext der übrigen Prädiktoren die erklärte Varianz in einer Weise erhöht, die man zunächst (aufgrund der einfachen Korrelationen) nicht vermutet hätte. Bortz (1977, p. 597) definiert als Suppressorvariablen solche, die den Vorhersagebeitrag einer (oder mehrerer) anderen Variablen erhöht, indem sie irrelevante Varianzen in der (den) anderen Prädiktorvariablen unterdrückt. Bortz operationalisiert dies in Anlehnung an Conger & Jackson (1972) und Conger (1974) für den Fall zweier Prädiktoren dadurch, daß er  $b^*_1 r_{y1} > r^2_{y1}$ , fordert, wenn  $x_2$  eine Suppressorvariable sein soll. Darlington (1968) polt alle Prädiktoren so um, daß ihre einfache Korrelation mit dem Kriterium nicht negativ ist, und nennt die Prädiktoren

Tabelle 3: Die Ergebnisse zum Beispiel aus Tab. 2 für verschiedene Werte der Prädiktoreninterkorrelation bei festen Korrelationen mit dem Kriterium  $r_{y1} = .463$  und  $r_{y2} = .612$

$r_{12}$	Prädiktoren-		Regressions-		Zusätzlicher		Bestimmt-		$R^2$ -Komponenten		Überlappung $R^2_{y(12)} - R^2_{y(1-2)} - R^2_{y(2-1)}$
	interkorrelation	interkorrelation	koeffizienten	Varianzbeitrag	heitsmaß	$b_1^*$	$b_2^*$	$r_{y1}b_1^*$	$r_{y2}b_2^*$		
-.400	.84	.95	.60	.76	.97			.39	.58		.38
.000	.46	.61	.21	.37	.59			.21	.37		.00
.683	.08	.56	.00	.16	.38			.04	.34		.21
.757	.00	.61	.00	.16	.37			.00	.37		.21
.950	-.1.21	1.77	.14	.30	.52			-.56	1.08		.07
.984	-4.39	4.93	.61	.77	.99			-2.03	3.02		-.40

Suppressorvariablen, deren Regressionskoeffizient  $b_k$  negativ ist. Für den Fall zweier Prädiktoren folgt daraus die Bedingung  $r_{y2}r_{1-y} < r_{y1}r_{2-y}$ , wenn  $x_2$  eine Suppressorvariable sein soll. Gaensslen & Schubö (1976, p. 117, 123) sprechen dann von einer Suppressorvariablen, wenn sie für sich mit dem Kriterium praktisch nicht korreliert, aber im Zusammenhang mit den übrigen Prädiktoren den insgesamt erklärten Varianzanteil erhöht. Im Beispiel nennen sie  $x_2$  eine Suppressorvariable, wenn  $r_{y(2-1)}^2$  groß gegenüber  $r_{y2}^2$  bzw. wenn  $|r_{y2}r_{12}|$  groß gegenüber  $|r_{y1}|$  ist. Cohen & Cohen (1975, p. 91) nennen  $x_2$  dann eine Suppressorvariable, wenn  $b_2^*$  außerhalb des Intervalls zwischen 0 und  $r_{y2}$  liegt. Velicer (1978), der Conger kritisiert, spricht ähnlich wie Gaensslen & Schubö (1976) dann von Suppression, wenn  $r_{y(2-1)}^2$  größer als  $r_{y2}^2$  ist. Holling (1980) schließlich verwendet als Kriterium  $r_{y(2-1)} > r_{y2}$  bzw. gleichwertig  $\hat{r}_{y(2-1)} > \hat{r}_{y2}$ , wenn  $x_2$  so gepolt ist, daß  $r_{y2}$  positiv ist.

Unter *Kollinearität* oder Multikollinearität versteht man, daß die Matrix  $X'X$  singulär oder fast singulär ist. Den Fall der exakten Kollinearität hatten wir durch Annahme A5 in Kap. 1.2 ausgeschlossen; er soll für das weitere ausgeschlossen bleiben. Eine Verbesserung der Schätzungen für die Regressionskoeffizienten war unter dem Titel Ridge-Regression in Kap. 1.6 vorgeschlagen worden. Hier sollen noch praktische Auswirkungen der Kollinearität ange deutet werden.

Stellt man einen gemeinsamen Konfidenzbereich aus den Daten des Beispiels für die beiden Regressionskoeffizienten für standardisierte Daten

$$(110) \quad \beta_1^* = \frac{s_1}{s_y} \beta_1, \quad \beta_2^* = \frac{s_2}{s_y} \beta_2$$

auf, so erhält man aus Gl. (22) unter Annahme des klassischen Modells die Ungleichung

$$(111) \quad \sum_{k=1}^K \sum_{l=1}^K r_{kl} (\beta_k^* - b_k^*) (\beta_l^* - b_l^*) \leq \frac{K(1-R^2)}{N-K-1} F_\alpha (K, N-K-1),$$

welche das Innere einer Ellipse mit Mittelpunkt  $(b_1^*, b_2^*)$  beschreibt. In Abb. 5 sind die Konfidenzbereiche zu  $r_{y1} = .463$ ,  $r_{y2} = .612$  und verschiedene, aus Tab. 3 ausgewählte Werte für  $r_{12}$  abgebildet. Wir wollen hier davon absehen, daß die Mittelpunkte der Ellipsen in Abhängigkeit von  $r_{12}$  variieren. Was jetzt vor allem interessieren soll, ist, daß die Ellipsen um so eher gestreckt sind, je größer der Betrag von  $r_{12}$  ist. Für  $r_{12}=0$  erhält man den Grenzfall eines Kreises als Konfidenzbereich, was davon herröhrt, daß die Schätzungen  $b_1^*$  und  $b_2^*$  unkorreliert sind. Wenn  $r_{12}$  hoch ist, ist auch die negative Kovarianz der Schätzungen hoch; dies bedeutet, daß bei zufälliger Überschätzung des einen Regressionskoeffizienten der andere Regressionskoeffizient eher unterschätzt wird. Jeder Regressionskoeffizient kann dann für sich nicht zuverlässig ge-

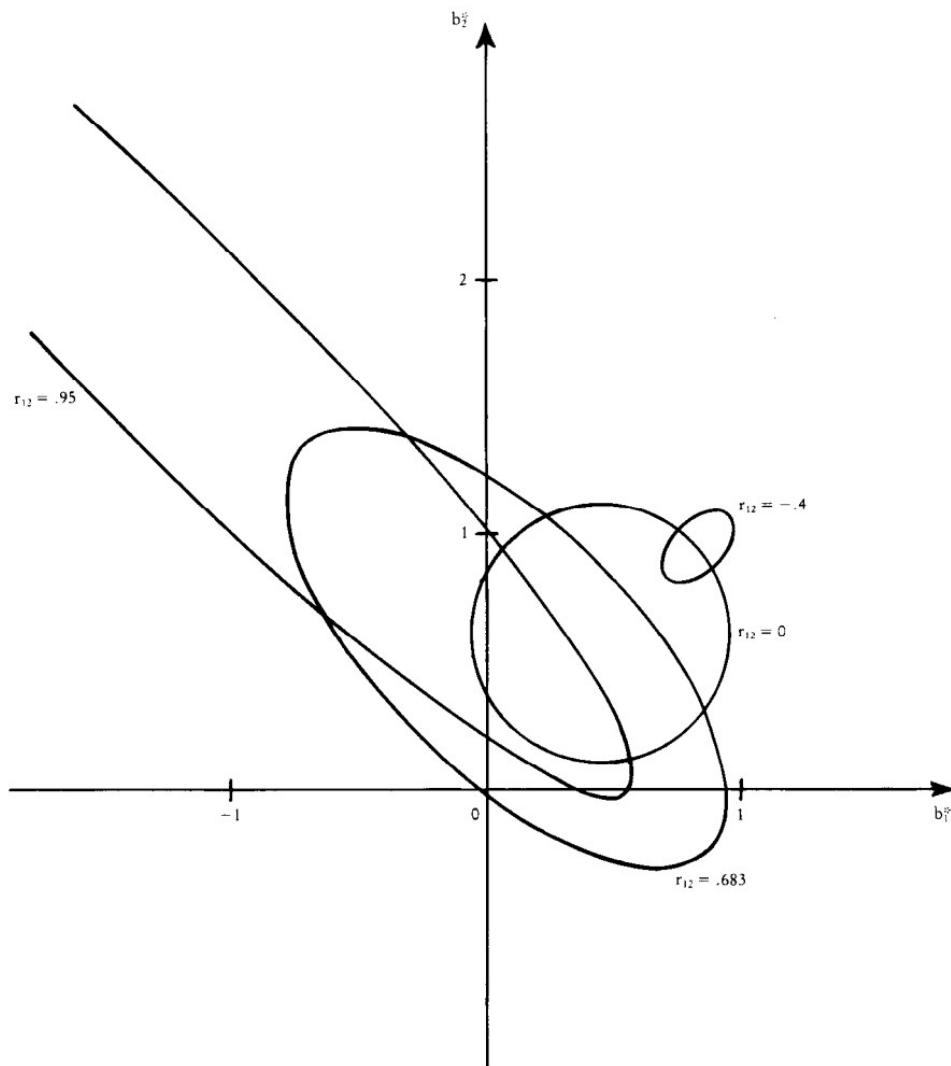


Abb. 5: Konfidenzbereiche der Regressionskoeffizienten für  $r_{y1} = .463$ ,  $r_{y2} = .612$  und verschiedene Werte der Prädiktoreninterkorrelation  $r_{12}$ .

schätzt werden. Die Zuverlässigkeit einer Prognose durch die Regressionsgleichung als Ganzes ist durch die Kollinearität der Prädiktoren jedoch nicht beeinträchtigt. über Möglichkeiten zur Orthogonalisierung von Prädiktoren berichtet Silvey (1969); Probleme der kollinearen Testauswertung diskutieren Schubö & Hentschel (1978) und Schubö (1980); über Schwierigkeiten bei Signifikanztests informiert Gordon (1968).

## 1.11 Schrittweise Regression

In vielen Bereichen der empirischen Forschung weiß man zwar, daß Zusammenhänge zwischen einem Kriterium und einer größeren Menge von Prädiktorvariablen bestehen, doch fällt eine Entscheidung darüber schwer, welche Untermenge der Prädiktorvariablen den Zusammenhang vermittelt. Eine Reduktion der Anzahl der Prädiktorvariablen ist aus Gründen der sparsameren Beschreibung wünschenswert. Wenn für das Forschungsgebiet keine Erfahrungen oder theoretische Hinweise vorliegen, wenn also keine *a priori* Reduktion der Prädiktoren möglich ist, ist man darauf angewiesen, die Variablenreduktion aufgrund der Zusammenhänge durchzuführen, die sich im vorliegenden Datensatz zeigen. Die in Kap. 1.10 angeführten Maße für die Wichtigkeit der Prädiktoren werden auch zu dem Zweck berechnet, Richtlinien für die Variablenreduktion zu erhalten. In diesem Kapitel sollen einige Methoden zur Auswahl der Prädiktoren diskutiert werden, die z.B. in Draper & Smith (1966, p. 163), Cohen & Cohen (1975, p. 102) und Graybill (1976, p. 439) vorgeschlagen werden.

Da die Potenzmenge der Menge der Prädiktoren endlich ist, steht im Prinzip dem nichts im Wege, der Reihe nach jede der  $2^k - 1$  möglichen nichtleeren Untermengen der Prädiktoren für Regressionsanalysen mit dem Kriterium zu verwenden. Diese Vorgehensweise wird als *Untermengen-Methode* (subset regression, all regressions method) bezeichnet und ist rechnerisch sehr aufwendig. Die „beste“ Einzelregression ist diejenige, die einerseits einen großen Varianzanteil erklärt, andererseits jedoch nur wenige Prädiktoren verwendet. Die Einzelregression mit dem besten Kompromiß zwischen diesen beiden gegensätzlichen Forderungen kann z.B. in einer ähnlichen Weise wie beim Scree-Test der Faktorenanalyse (z.B. Gaensslen & Schubö, 1976, p. 226) dadurch bestimmt werden, daß in einem Kreuzdiagramm das maximale Bestimmtheitsmaß für eine feste Zahl  $k$  von Prädiktoren gegen die Zahl  $k$  aufgetragen wird. Andere Kriterien und schnelle Algorithmen findet man bei Garside (1965), Hocking & Leslie (1967), LaMotte & Hocking (1970), Furnival (1971), Hocking (1972), Morgan & Tatar (1972) oder auch in IMSL (1979).

In sozialwissenschaftlichen Anwendungen ist allerdings für die Einschätzung der gefundenen besten Regression größte Vorsicht erforderlich. Eine Kreuzvalidierungsstudie zur Überprüfung der Regression (s. Kap. 1.5: Prognosetest) mit einer unabhängigen Stichprobe ist obligat. Es genügt keinesfalls, für die gefundene Regression einen Signifikanztest durchzuführen, weil die tatsächliche Irrtumswahrscheinlichkeit erheblich das gewählte Signifikanzniveau übersteigt. Aus Schubö (1982) wurde Tab. 4 entnommen, die für das Signifikanzniveau 5% die effektive Irrtumswahrscheinlichkeit angibt. Beispielsweise ist die Wahrscheinlichkeit dafür, daß der Zusammenhang zwischen den drei besten Prädiktoren aus insgesamt 13 Prädiktoren auf dem 5%-Niveau signifikant

wird, gleich 58%, wenn in Wirklichkeit keinerlei Zusammenhang zwischen dem Kriterium und den Prädiktoren und zwischen den Prädiktoren untereinander besteht.

Tabelle 4: Gesamt-Irrtumswahrscheinlichkeit bei der empirischen Auswahl der Prädiktoren mit dem stärksten Zusammenhang zum Kriterium (Irrtumswahrscheinlichkeit  $\alpha = .05$  für den Einzeltest).

Gesamtzahl der Prädik- toren K	Anzahl verwendeter Prädiktoren	1	2	3	4	5	6	7	8	9	10	beliebig
2	.10	.05	-	-	-	-	-	-	-	-	-	.12
3	.14	.10	.05	-	-	-	-	-	-	-	-	.17
4	.19	.14	.09	.05	-	-	-	-	-	-	-	.22
5	.23	.19	.14	.08	.05	-	-	-	-	-	-	.27
6	.26	.26	.19	.14	.09	.05	-	-	-	-	-	.32
7	.30	.31	.24	.17	.13	.08	.05	-	-	-	-	.38
8	.34	.36	.32	.25	.18	.13	.08	.05	-	-	-	.43
9	.37	.39	.36	.30	.24	.17	.12	.08	.05	-	-	.46
10	.40	.43	.41	.37	.31	.23	.18	.12	.08	.05	.05	.51
11	.43	.50	.46	.42	.36	.29	.22	.16	.12	.09	.09	.56
12	.46	.53	.52	.49	.44	.37	.30	.23	.17	.12	.12	.61
13	.49	.56	.58	.55	.51	.45	.38	.30	.23	.17	.17	.66
14	.51	.60	.62	.61	.56	.51	.43	.36	.29	.22	.22	.72
15	.54	.64	.67	.65	.61	.56	.49	.43	.36	.29	.29	.75
16	.56	.68	.71	.70	.67	.63	.58	.52	.43	.35	.35	.77
17	.58	.69	.72	.73	.71	.67	.62	.57	.49	.43	.43	.80
18	.60	.73	.77	.76	.75	.71	.67	.61	.56	.50	.50	.83
19	.62	.76	.80	.80	.79	.75	.72	.67	.62	.55	.55	.85
20	.64	.78	.82	.83	.82	.81	.78	.75	.70	.64	.64	.88
25	.72	.84	.90	.91	.91	.89	.87	.85	.83	.80	.80	.96
29	.77	.91	.93	.96	.97	.96	.95	.94	.93	.92	.92	.98

Dieselbe Problematik besteht bei drei weiteren Methoden zur empirischen Auswahl von Prädiktoren, der Vorwärts-Methode, der Rückwärts-Methode und der Austausch-Methode. Alle diese Methoden sollen mit geringerem Aufwand eine beinahe so gute Untermenge der Prädiktoren finden, wie sie die Untermengen-Methode liefert. Bei allen Methoden werden schrittweise Prädiktoren in die Regressionsgleichung aufgenommen bzw. wieder ausgeschieden. Keine der drei Methoden führt mit Sicherheit zur definiert besten Untermenge der Prädiktoren im zuvor genannten Sinn.

Bei der *Vorwärts-Methode* wird zunächst die einzelne, mit dem Kriterium am stärksten korrelierende Variable aufgenommen, dann als zweiter Prädiktor die Variable, die den größten Zugewinn des Zusammenhangs bewirkt, usw. Bei der *Rückwärts-Methode* werden aus dem Regressionsansatz mit allen Prädiktoren jeweils die Variablen ausgeschieden, deren Entfernen die kleinste Veränderung des Zusammenhangs zur Folge hat. Bei der *Austausch-Methode* wird im Prinzip wie bei der Vorwärts-Methode vorgegangen, doch wird nach der Aufnahme neuer Prädiktoren in den Regressionsansatz jeweils geprüft, ob andere Prädiktoren aus ihm entfernt werden können, ohne daß der Gesamtzusammenhang erheblich zurückgeht. Gleichzeitig kann bei allen Methoden die Aufnahme näherungsweise kollinearer Prädiktoren ausgeschlossen werden. Für den endgültigen Aufbau der Regressionsgleichung ist es im Fall der Kollinearität zweier oder mehrerer Prädiktoren meist dem Zufall vorbehalten, welche Prädiktoren herangezogen werden. Beinahe alle Computerprogramme ermöglichen die Prädiktorenauswahl nach wenigstens einer dieser drei Methoden, wobei leider für die endgültige Prädiktorenauswahl der Signifikanztest mit falscher Irrtumswahrscheinlichkeit berechnet wird. Der Anwender dieser Programme wird zwar subjektiv glücklicher, weil er einen signifikanten Zusammenhang berichten kann, doch besteht sein unbemerktes Unglück darin, daß er mit aller Wahrscheinlichkeit eine nicht gerechtfertigte Entscheidung trifft.

Eine erheblich geringere Verfälschung der Irrtumswahrscheinlichkeit ist durch eine Vorgehensweise für Prädiktorengruppen gegeben, wie sie von Cohen & Cohen (1975, p. 123) ausführlich beschrieben wird. Dabei werden die Prädiktoren *a priori* in disjunkte Variablengruppen A, B, usw. eingeteilt. Die Einteilung erfolgt entweder nach strukturellen oder nach funktionalen Gesichtspunkten. Ein struktureller Gesichtspunkt ist etwa, wenn die  $K_A = g-1$  Kodiervariablen (z.B. Dummyvariablen) zu einer qualitativen Variablen mit  $g$  Ausprägungen in eine Gruppe A zusammengefaßt wird, oder wenn nichtlineare Effekte einer quantitativen, standardnormalverteilten Variablen  $z$  durch eine Gruppe B von  $K_B = 3$  orthogonalen Hermiteschen Polynomen  $z$ ,  $z^2-1$ ,  $z^3-3z$  berücksichtigt werden. Nach funktionalen Gesichtspunkten könnte man Prädiktoren des Studienerfolgs etwa in die Gruppen Reifezeugnisnoten (A), Leistungstest-variablen (B), motivationale Variablen (C) und Lebenslauf-Daten (D) einteilen.

Um die Vorgehensweise übersichtlich darstellen zu können, müssen wir zunächst eine kompakte Schreibweise für die Behandlung von Variablengruppen einführen. Wenn die Gruppe A von Prädiktorvariablen

$$A = \{ x_1, x_2, \dots, x_{K_A} \}$$

ist, schreibt man für die multiple Korrelation kurz

$$R_{y(A)} = R_{y(12 \dots K_A)}$$

und, wenn C die Vereinigung von A und B ist,

$$R_{y(AB)} = R_{y(C)}.$$

Die quadrierte semipartiale multiple Korrelation

$$(112) \quad R_{y(A-B)}^2 = R_{y(AB)}^2 - R_{y(B)}^2$$

ist dann der Varianzanteil des Kriteriums y, der durch die Hinzunahme von A zusätzlich zu B erklärt wird. Für den Fall dreier Gruppen von Prädiktoren kann in dieser Begriffsbildung eine hierarchische Zerlegung

$$(113) \quad R_{y(ABC)}^2 = R_{y(A)}^2 + R_{y(B-A)}^2 + R_{y(C-AB)}^2$$

der erklärten Varianz vorgenommen werden. Ob die zu den Variablen der Gruppe A gehörenden Regressionskoeffizienten alle Null sind

$$\beta_1 = \beta_2 = \dots = \beta_{K_A} = 0,$$

kann durch die Entscheidungsregel zum Verwerfen dieser Nullhypothese

$$(114) \quad \frac{(N-K_A-K_B-1) (R_{y(AB)}^2 - R_{y(B)}^2)}{K_A(1-R_{y(AB)}^2)} > F_{\alpha}(K_A, N-K_A-K_B-1)$$

entschieden werden, wenn die Annahmen des klassischen Modells für den Ansatz mit den zwei Sätzen A und B von Prädiktoren erfüllt sind. Für den Fall, daß A nur eine Variable enthält, entspricht dies genau der Entscheidungsregel (37).

Welche Strategien sich für die Signifikanztests bei Variablengruppen anbieten, soll nun am bereits erwähnten Beispiel der Kriteriumsvariablen Studienerfolg demonstriert werden. Einerseits kann im Rahmen des vollen Satzes von Prädiktoren ABCD jeweils geprüft werden, ob die einzelnen Gruppen zusätzlich zu den drei übrigen Gruppen signifikante Varianzanteile erklären. Beispielsweise wäre der Test dafür, ob die motivationalen Variablen C zusätzlich zu den Reifezeugnis-Noten A, den Leistungstest-Variablen B und den Lebenslauf-Daten D Varianzanteile erklären, durch

$$(115) \quad \frac{(N-K-1) (R_{y(ABCD)}^2 - R_{y(ABC)}^2)}{K_C(1-R_{y(ABCD)}^2)} > F_{\alpha}(K_C, N-K-1)$$

gegeben, wobei die Gesamtzahl der Prädiktoren

$$K = K_A + K_B + K_C + K_D$$

ist. Diese erste Strategie entspricht der ersten Stufe der Rückwärtsmethode.

Entsprechend der Vorwärtsmethode wird die zweite Strategie gewählt, mit dem Unterschied, daß nach inhaltlichen Gesichtspunkten a priori eine Hierarchie für die Aufnahme in die Regression festgelegt wird. Beispielsweise könnte es vernünftig erscheinen zu untersuchen, ob erstens die Reifezeugnisnoten A für sich mit dem Studienerfolg zusammenhängen, ob zweitens die Leistungstest-Variablen B eine Verbesserung der Stärke des Zusammenhangs mit sich bringen, ob drittens die motivationalen Variablen C zusätzlich zu A und B beitragen, und ob schließlich die Aufnahme von Lebenslauf-Daten D sinnvoll oder erforderlich ist. Wieder sei die Entscheidungsregel für die Aufnahme der motivationalen Variablen C als Beispiel angeführt:

$$(116) \quad \frac{(N - K_A - K_B - K_C - 1) (R^2_{y(ABC)} - R^2_{y(AB)})}{K_C (1 - R^2_{y(ABC)})} > F_{\alpha}(K_c, N - K_A - K_B - K_C - 1)$$

Bei beiden Vorgehensweisen werden zunächst nur vier Signifikanztests durchgeführt, wodurch sich die Gesamt-Irrtumswahrscheinlichkeit nicht in dem Maße wie bei der Untermengen-, Vorwärts-, Rückwärts- oder Austausch-Methode erhöht. Wenn man es wünscht, können weitere Signifikanztests in Form von geschützten Tests innerhalb der „signifikanten“ Variablengruppen für die einzelnen Variablen angeschlossen werden. Man kann dabei hoffen, daß die endgültige Variablenauswahl nicht in dem Maße zufallsabhängig ist, wie dies bei den zuerst genannten Auswahlstrategien der Fall ist. Doch stehen zuverlässige Untersuchungen dazu noch aus.

## 1.12 Teststärke

Bisher haben wir im Zusammenhang mit der Prüfung statistischer Hypothesen immer nur von der Wahrscheinlichkeit, einen Fehler 1. Art zu begehen, gesprochen. Beim Prüfen statistischer Hypothesen können jedoch grundsätzlich zwei Fehlentscheidungen getroffen werden. Einmal kann eine tatsächlich richtige Hypothese aufgrund eines Tests verworfen werden, andererseits kann eine tatsächlich falsche Hypothese nicht verworfen werden. Solange man jedoch nur eine einzige Hypothese betrachtet, lassen sich nur Aussagen über die Wahrscheinlichkeit des Fehlers 1. Art, das vorzugebende Signifikanzniveau  $\alpha$ , machen. Dies ist auch in dem von Fisher ursprünglich konzipierten Signifikanztest die einzige bestimmende Größe (vgl. Haagen & Seifert, 1979, Kap. 9, insbesondere p. 198-199). Solange man jedoch nur eine einzige Hypothese prüft, lassen sich keine Aussagen über die Wahrscheinlichkeit eines Fehlers 2. Art machen, und damit auch keine Aussagen über die „Güte“ eines Tests, wenn man die Güte eines Tests durch die Wahrscheinlichkeit, richtig zu entscheiden, mißt. Zur Beurteilung der Güte eines Tests ist es also notwendig, neben der Nullhypothese auch die Alternativhypothese zu spezifizieren, wie dies auch im Rahmen der Neyman-Pearson-Testtheorie geschieht. Die *Güte*

*funktion (Teststärke, Trennschärfe)* eines Tests gibt für den Bereich der Alternativhypothese die Wahrscheinlichkeit dafür an, die Nullhypothese zu verwerfen, d.h. bei Gültigkeit der Alternativhypothese eine richtige Entscheidung zu treffen (vgl. Haagen & Seifert, 1979, p. 209). Bezeichnen wir mit  $\beta^1$  die Wahrscheinlichkeit, einen Fehler 2. Art zu begehen, so ergibt sich für die Teststärke 1 -  $\beta$ . Beim Prüfen statistischer Hypothesen ist man nun speziell an solchen Tests interessiert, die bei gegebenem Signifikanzniveau eine möglichst große Teststärke aufweisen (vgl. dazu auch die Ausführungen in Kap. 1.5).

Um die Teststärke bei der Prüfung der multiplen Korrelation zu berechnen, müssen wir also neben dem Stichprobenumfang N und der Zahl der Prädiktoren K noch die Größe der multiplen Korrelation R in der Grundgesamtheit festlegen. Durch diese Größen ist über die Verteilungsannahme A5 die Verteilung der Prüfgröße F, die nichtzentrale F-Verteilung (siehe z.B. Tang, 1938; Laubscher, 1960) vollständig bestimmt. Man nennt R oder meist die daraus abgeleitete Größe

$$f^2 = \frac{R^2}{1-R^2}$$

auch die Effektgröße. Im allgemeinen wird man die Effektgröße allerdings nicht kennen. Doch hat man häufig aufgrund von Ergebnissen ähnlicher Studien Vorstellungen darüber, in welchem Intervall die Effektgröße liegen wird. Sicherlich wird es jedenfalls möglich sein, eine Mindest-Effektgröße anzugeben, bei der man noch von einem bedeutsamen Effekt oder Zusammenhang sprechen könnte. In der Persönlichkeitsforschung etwa könnte man als absolutes Minimum von bedeutsamen Zusammenhängen eine Korrelation  $R = .2$  (4% erklärte Varianz), also eine Effektstärke  $f^2 = .042$  annehmen.

Allgemein läßt sich sagen, daß die Teststärke bei der Prüfung der multiplen Korrelation um so größer wird,

- a) je größer das Signifikanzniveau  $\alpha$  ist,
- b) je größer der Stichprobenumfang N ist,
- c) je größer die Effektstärke ist und
- d) je kleiner die Anzahl der Prädiktoren ist.

<sup>1)</sup> Da die Wahrscheinlichkeit für einen Fehler 2. Art unter sonst gleichen Umständen von den Parametern des Parameterraumes der Alternativhypothese (es werden hier nur parametrisierbare Wahrscheinlichkeitsverteilungen (vgl. Haagen & Seifert, 1979, p. 206) betrachtet), müßten wir genauer  $\beta$  als abhängig von diesen Parametern symbolisieren. Aus schreibtechnischen Gründen verzichten wir jedoch darauf.

Diese Einflußgrößen hängen in der Regel von äußeren Bedingungen ab; am ehesten scheinen noch der Stichprobenumfang und die Zahl der Prädiktoren relativ frei wählbar zu sein.

Ohne ausreichende Teststärke besteht kaum Aussicht, eine Entscheidung für oder gegen die Nullhypothese zu fallen. Bei der Planung einer Untersuchung sollte daher eine Abschätzung der Teststärke Pflicht sein. Als eine Art Standard kann eine Teststärke von .80 angesehen werden.

Eine ausführliche Diskussion der Teststärken aller im Zusammenhang mit der multiplen Regression üblichen Tests findet sich in Cohen & Cohen (1975). Mit noch besserer Genauigkeit werden Teststärken in Laubscher (1969), Lehmer (1944) und Tang (1938) angegeben.

### *1.12.1 Teststärke für die Prüfung der multiplen Regression*

Anstatt für vorgegebenes Signifikanzniveau  $\alpha$ , Stichprobenumfang  $N$ , Effektstärke  $f^2$  und Anzahl der Prädiktoren  $K$  die Teststärke  $1 - \beta$  zu bestimmen, soll hier von einer vorgegebenen Teststärke und dem vorgegebenen Signifikanzniveau  $\alpha$  ausgegangen werden; sind zwei der übrigen drei Größen gegeben, kann dann die dritte so bestimmt werden, daß der Test gerade mit der vorgegebenen Teststärke durchgeführt werden kann.

### *1.12.2 Bestimmen des erforderlichen Stichprobenumfangs $N$*

Der Stichprobenumfang kann mit der Näherungsformel

$$(117) \quad N = \frac{L}{f^2} + K + 1$$

berechnet werden, wobei

$$(118) \quad L = c + \sqrt{dK - e}$$

in den in Tab. 5 aufgeführten Koeffizienten die Abhängigkeit vom Signifikanzniveau  $\alpha$  und der Teststärke  $1 - \beta$  enthält und die Effektstärke

$$(119) \quad f^2 = \frac{R^2}{1 - R^2}$$

von der vermuteten Populationskorrelation abhängt. Zusammengefaßt ergibt sich

$$(120) \quad N = \frac{1 - R^2}{R^2} (c + \sqrt{dK - e}) + K + 1.$$

Tabelle 5: Koeffizienten zur näherungsweisen Berechnung der L-Werte

Teststärke $1 - \beta$	Signifikanzniveau $\alpha = .05$			Signifikanzniveau $\alpha = .01$		
	c	d	e	c	d	e
.10	- . 1 6	.26	- . 0 9	.16	2.12	- . 2 9
.30	.65	2.47	.42	1.95	6.38	1.02
.50	1.95	5.39	1.66	3.92	10.72	3.19
.60	2.87	7.17	3.08	5.09	13.23	4.66
.70	3.95	9.40	4.67	6.48	16.24	6.52
.75	4.61	10.78	5.60	7.31	18.05	7.60
.80	5.41	12.43	6.82	8.30	20.18	9.02
.85	6.52	14.37	9.20	9.52	22.81	10.90
.90	7.79	17.34	10.77	11.17	26.37	13.40
.95	10.05	22.07	14.69	13.85	32.15	17.83
.99	15.04	32.55	24.01	19.58	44.68	27.50

Obwohl die Auswertung dieser Formel im Einzelfall nicht schwierig ist, wurden die Ergebnisse für den vielleicht wichtigsten Fall  $\alpha = .05$  und  $1 - \beta = .80$  in Tab. 6 mit etwas besserer Genauigkeit (relativer Fehler kleiner 1%) aufgenommen. Im Falle der einfachen Korrelation ( $K = 1$ ) ergaben sich allerdings aus den Tabellen G.2 und E.2 bei Cohen & Cohen (1975) und ebenso aus den Tabellen 3.4.1 und 8.4.4 bei Cohen (1969) Abweichungen, die etwas über den von den Autoren angegebenen Fehlergrenzen liegen; unsere Tab. 6 enthält die jeweils größeren Forderungen für den Stichprobenumfang.

Die Näherung (117) stammt von Cohen & Cohen (1975, p. 118). Die Beziehung (118) ist eine Approximation, für die die in Tab. 5 aufgeführten Koeffizienten durch eine Regressionsanalyse mit dem Kriterium  $K$  und den Prädiktoren  $L$  und  $L^2$  aus den Tabellen von Cohen & Cohen (1975) bestimmt wurden. Um den Approximationsfehler für kleine Werte von  $K$  niedrig zu halten, wurde für  $K = 1$  zehnfach und für  $K = 2$  bis 5 jeweils fünffach gewichtet. Der relative Fehler der Näherung (118) gegenüber den tabellierten Werten von  $L$  beträgt höchstens 1% mit Ausnahme des Falles  $K = 1$ , wo relative Fehler bis zu 2,5% auftreten (in Tab. E2 bei Cohen & Cohen, 1975, ist für  $1 - \beta = .50$  und  $K = 24$  ein Druckfehler enthalten: statt 13.02 muß  $L = 13.25$  stehen). Diese nicht sehr erheblichen numerischen Fehler werden dadurch aufgewogen, daß es durch die Näherung (118) möglich wird, in Gl. (120) eine algebraische Beziehung zwischen dem Stichprobenumfang  $N$ , der Populationskorrelation  $R$  und der Prädiktorenzahl  $K$  für eine bestimmte Irrtumswahrscheinlichkeit und Teststärke aufzustellen. Diese Beziehung kann nach jeder der drei Größen aufgelöst werden, was im folgenden Verwendung findet.

Tabelle 6: Mindestens erforderlicher Stichprobenumfang N in Abhängigkeit von der Anzahl der Prädiktoren und der Größe der multiplen Korrelation in der Grundgesamtheit für den Signifikanztest der multiplen Korrelation aus der Stichprobe mit einem Signifikanzniveau von  $\alpha = 0.05$  und einer Teststärke  $1 - \beta = 0.80$

Anzahl der Prädik- toren	Multiple Korrelation in der Grundgesamtheit												
	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99
1	78812	3145	783	193	84	46	28	18	12	9	6	4	3
2	96544	3856	959	235	101	54	32	21	14	9	6	5	4
3	109269	4365	1086	267	115	62	37	24	16	11	7	6	5
4	119567	4776	1189	292	126	68	41	27	18	12	8	7	6
5	128453	5132	1278	315	136	74	45	29	20	14	10	8	7
6	136385	5450	1358	335	145	79	48	32	22	15	11	9	8
7	143618	5739	1430	353	154	84	52	34	23	17	12	10	9
8	150308	6007	1498	370	161	88	55	36	25	18	13	11	10
9	156563	6258	1561	386	169	93	57	38	27	19	14	12	11
10	162457	6494	1620	401	176	97	60	40	28	21	15	13	12
11	168047	6718	1676	416	182	101	63	42	30	22	16	14	13
12	173376	6931	1730	430	189	105	66	44	32	23	18	15	14
14	183376	7332	1831	456	201	112	71	48	35	26	20	17	16
16	192656	7705	1925	480	212	119	75	52	38	28	22	20	18
18	201354	8054	2013	503	223	125	80	55	40	31	24	22	20
20	209565	8383	2096	524	233	132	84	59	43	33	26	24	22
25	228411	9140	2288	575	257	146	95	67	50	39	32	29	27
30	245409	9823	2461	620	280	160	105	75	57	45	37	34	32
35	261017	10451	2620	663	300	174	115	83	64	51	43	39	37
40	275528	11035	2769	703	320	186	124	90	70	57	48	44	42
45	289146	11583	2909	740	339	198	133	98	77	63	53	50	47
50	302017	12101	3041	776	357	210	142	105	83	68	59	55	52
60	325938	13065	3288	844	391	233	159	119	95	80	69	65	62
70	347919	13952	3516	906	423	254	176	133	108	91	80	75	72
80	368368	14778	3728	965	454	275	192	147	120	102	90	85	82
90	387566	15553	3928	1022	483	295	208	160	132	113	101	96	92
100	405718	16287	4118	1075	512	314	223	174	144	124	111	106	102

Tabelle 7: Erforderliche Populationskorrelation R für einen Signifikanztest mit Signifikanzniveau  $\alpha = .05$ . und Teststärke  $1 - \beta = .80$  in Abhängigkeit vom Stichprobenumfang N und der Anzahl der Prädiktoren K.

Stich- proben- umfang N	Anzahl der Prädiktoren K															
	1	2	3	4	5	6	7	8	9	10	15	20	30	50	75	100
3	.99	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
4	.96	.96	-	-	-	-	-	-	-	-	-	-	-	-	-	
5	.92	.92	.96	-	-	-	-	-	-	-	-	-	-	-	-	
6	.88	.88	.92	.97	-	-	-	-	-	-	-	-	-	-	-	
7	.85	.85	.89	.93	.97	-	-	-	-	-	-	-	-	-	-	
8	.82	.82	.86	.90	.94	.97	-	-	-	-	-	-	-	-	-	
9	.79	.79	.83	.87	.91	.94	.97	-	-	-	-	-	-	-	-	
10	.77	.77	.81	.84	.88	.91	.94	.97	-	-	-	-	-	-	-	
12	.71	.72	.76	.80	.83	.86	.89	.92	.95	.98	-	-	-	-	-	
14	.67	.69	.73	.76	.79	.82	.84	.87	.90	.92	-	-	-	-	-	
16	.64	.66	.70	.73	.75	.78	.81	.83	.86	.88	-	-	-	-	-	
18	.61	.63	.67	.70	.72	.75	.77	.80	.82	.84	.96	-	-	-	-	
20	.58	.61	.64	.67	.70	.72	.74	.76	.79	.81	.91	-	-	-	-	
25	.53	.56	.59	.62	.64	.66	.68	.70	.72	.74	.83	.92	-	-	-	
30	.49	.52	.55	.57	.60	.62	.63	.65	.67	.68	.76	.84	-	-	-	
35	.45	.49	.52	.54	.56	.58	.59	.61	.63	.64	.71	.78	.93	-	-	
40	.43	.46	.49	.51	.53	.55	.56	.58	.59	.60	.67	.73	.86	-	-	
45	.40	.44	.46	.48	.50	.52	.53	.55	.56	.57	.63	.69	.80	-	-	
50	.38	.42	.44	.46	.48	.50	.51	.52	.54	.55	.60	.65	.76	-	-	
60	.35	.39	.41	.43	.44	.46	.47	.48	.49	.50	.55	.60	.68	.88	-	
70	.33	.36	.38	.40	.41	.43	.44	.45	.46	.47	.51	.55	.63	.79	-	
80	.31	.34	.36	.38	.39	.40	.41	.42	.43	.44	.48	.52	.58	.72	.95	
100	.28	.31	.32	.34	.35	.36	.37	.38	.39	.40	.43	.46	.52	.62	.78	
125	.25	.28	.29	.31	.32	.33	.34	.34	.35	.36	.39	.41	.46	.54	.65	.80
150	.23	.25	.27	.28	.29	.30	.31	.32	.32	.33	.36	.38	.42	.49	.58	.68
175	.21	.24	.25	.26	.27	.28	.29	.29	.30	.31	.33	.35	.39	.45	.52	.60
200	.20	.22	.23	.25	.25	.26	.27	.28	.28	.29	.31	.33	.36	.42	.48	.54
300	.16	.18	.19	.20	.21	.22	.22	.23	.23	.24	.25	.27	.29	.33	.38	.42
400	.14	.16	.17	.18	.18	.19	.19	.20	.20	.21	.22	.23	.25	.29	.32	.35
500	.13	.14	.15	.16	.16	.17	.17	.18	.18	.18	.20	.21	.23	.26	.28	.31
750	.11	.12	.13	.13	.14	.14	.14	.15	.15	.15	.16	.17	.19	.21	.23	.25
1000	.09	.10	.11	.11	.12	.12	.12	.13	.13	.13	.14	.15	.16	.18	.20	.21
2000	.07	.07	.08	.08	.09	.09	.09	.09	.09	.10	.10	.11	.12	.13	.14	.15

### 1.12.3 Die erforderliche Populationskorrelation $R$

Sind Signifikanzniveau  $\alpha$ , Teststärke  $1 - \beta$ , Anzahl der Prädiktoren  $K$  und der Stichprobenumfang  $N$  gegeben, dann kann mit folgender Umformung der Gl. (120)

$$(121) \quad R = 1 / \sqrt{\frac{N - K - 1}{c + \sqrt{dK - e}}} + 1$$

die für diesen Signifikanztest mindestens erforderliche Größe der multiplen Korrelation in der Grundgesamtheit bestimmt werden. Die Ergebnisse für  $\alpha = .05$  und  $1 - \beta = .80$  werden in Tab. 7 dargestellt, wobei wiederum für  $K = 1$  abweichend von Gl. (121) die ungünstigsten Angaben aus Cohen (1969) herangezogen wurden.

### 1.12.4 Die höchstens sinnvolle Prädiktorenanzahl $K$

In der Praxis tritt schließlich häufig das Problem auf, bei bekannter Abschätzung der Stärke eines Zusammenhangs zwischen einem Kriterium und einem Satz von Prädiktorvariablen, die höchstens sinnvolle Anzahl von Prädiktorvariablen festzulegen. Löst man Gl. (120) nach  $K$  auf, erhält man

$$(122) \quad K = N - 1 - f^2 (c + \sqrt{f^4(d/2)^2 + d(N-1) - f^2cd - e} - f^2(d/2))$$

$$\text{mit } f^2 = \frac{R^2}{1 - R^2}.$$

Auch für diesen Fall werden die Ergebnisse in Tab. 8 dargestellt, wobei für  $K = 1$  die besprochene Korrektur wieder vorgenommen wurde. Hier wurden die Werte stets abgerundet, während sie bei den Tabellen 6 und 7 stets nach oben aufgerundet wurden, damit die behauptete Teststärke auf alle Fälle erreicht wird.

### 1.12.5 Teststärke für die übrigen Signifikanztests bei der Regressionsanalyse

Bisher wurde nur berichtet, wie die Teststärke für den Signifikanztest der multiplen Korrelation, d.h. für die Nullhypothese  $\beta_1 = \beta_2 \dots = \beta_K = 0$ , berechnet werden kann. Die Teststärke aller übrigen Signifikanztests, die sich auf eine unter der Nullhypothese F-verteilte Prüfgröße stützen, erhält man auf entsprechende Weise. Auch bei diesen Signifikanztests gehört die Verteilung der Prüfgröße F unter einer alternativen Hypothese zur Klasse der nichtzentralen F-Verteilungen. Entsprechend können die Gleichungen (117), (118) und (119) in etwas abgeänderter Form eingesetzt werden.

Tabelle 8: Größte Anzahl K von Prädiktoren, für die die Teststärke 1 -  $\beta$  mindestens .80 beträgt bei einem Signifikanzniveau des Signifikanztests für die multiple Korrelation von  $\alpha = .05$ .

(Wenn die Teststärke .80 stets unterschritten wird, wird dies durch - symbolisiert; das Symbol + steht, wenn bei 100 Prädiktoren die Teststärke .80 überschritten wird.)

Für den Signifikanztest eines einzelnen Regressionskoeffizienten  $b_1$  (Nullhypothese  $\beta_1 = 0$ , zweiseitige Alternativhypothese), der zugleich der Signifikanztest der partiellen Korrelation  $r_{y(1-23 \dots K)}$  und der semipartiellem Korrelation  $r_{y(l-23 \dots K)}$  ist, wird nach Cohen & Cohen (1975) als Effektstärke die für die Gesamtheit abgeschätzte Größe

$$(123) \quad f^2 = \frac{r_{y(1-23 \dots K)}^2}{1 - R^2} = \frac{R_{y(12 \dots K)}^2 - R_{y(23 \dots K)}^2}{1 - R_{y(12 \dots K)}^2}$$

verwendet. Gleichung (117) bleibt unverändert, doch muß in Gl. (118) der Wert 1 für K eingesetzt werden:

$$(124) \quad L = c + \sqrt{d - e}.$$

Soll der zusätzliche Beitrag einer Variablengruppe B zur Schätzung des Kriteriums y, zusätzlich zu einer anderen Variablengruppe A, geprüft werden, so entspricht dies der Prüfung der Nullhypothese, daß alle Regressionskoeffizienten der Variablen der Gruppe B in der Grundgesamtheit Null sind. Dies ist gleichzeitig der Test für die multiple semipartiellem Korrelation

$$(125) \quad R_{y(B-A)}^2 = R_{y(AB)}^2 - R_{y(A)}^2.$$

Dementsprechend wird die für die Alternativhypothese postulierte Effektstärke mit

$$(126) \quad f^2 = \frac{R_{y(AB)}^2 - R_{y(A)}^2}{1 - R_{y(AB)}^2}$$

bestimmt. Die Gleichungen (117) und (118) müssen abgeändert werden zu

$$(127) \quad N = \frac{L}{f^2} + K_{AB} + 1$$

und

$$(128) \quad L = c + \sqrt{d K_B - e}$$

wobei  $K_B$  die Zahl der Variablen der Gruppe B und  $K_{AB}$  die Gesamtzahl der Variablen in beiden Gruppen A und B sind. In Gleichung (128) wird stets als Wert K die Zahl der Freiheitsgrade für den Zähler von Gleichung (126) eingesetzt, hingegen in Gleichung (127) die Zahl der Variablen, die bei der Bildung des Fehlervarianz-Anteils im Nenner von Gleichung (126) beteiligt sind.

Diese Art der Bestimmung der Teststärke ist nicht auf quantitativ gemessene Prädiktorvariablen beschränkt, sie kann ebenso für Prädiktoren erfolgen, durch die Kategorien von qualitativen Variablen dargestellt werden. So kann auch für die univariate Varianzanalyse mit dem selben Formalismus die Teststärke bzw. der erforderliche Stichprobenumfang für eine bestimmte Teststärke angegeben werden. Wird die Varianzanalyse nicht über die Nomenklatur

der multiplen Regression sondern mit der traditionell auf Mittelwertsdifferenzen aufbauenden Symbolik beschrieben, dann tritt lediglich an die Stelle der quadrierten multiplen Korrelation  $R^2$  das Verhältnis  $\eta^2$  der Varianz zwischen den Gruppen zur Gesamtvarianz.

## 2. Kanonische Korrelation

### 2.1 Einführung

Die kanonische Korrelation kann als Verallgemeinerung der multiplen Regression im korrelationsanalytischen Modell (vgl. Kap. 1.7) gesehen werden. Bei der multiplen Regression besteht das Problem darin, eine Kriteriumsvariable mit Hilfe von zwei oder mehr Prädiktorvariablen linear zu erklären, so daß eine Funktion des Vorhersagefehlers möglichst klein wird. Dies ist äquivalent damit, die Koeffizienten der Linearkombination der Prädiktorvariablen so zu bestimmen, daß die Korrelation zwischen der Kriteriumsvariable und der Linearkombination der Prädiktorvariablen möglichst groß ist. Die kanonische Korrelation stellt insofern eine Verallgemeinerung der multiplen Regression dar, als gleichzeitig mehr als eine Kriteriumsvariable in die Analyse einbezogen wird.

Allgemein läßt sich die Problemstellung der kanonischen Korrelation folgendermaßen beschreiben: Gegeben sind eine Menge von mindestens zwei Kriteriumsvariablen und eine Menge von Prädiktorvariablen, die ebenfalls zwei oder mehr Variablen enthält. Ziel ist es, den Zusammenhang zwischen den beiden Variablenmengen mit Hilfe von möglichst wenigen und möglichst einfachen aus den beiden Variablenmengen abgeleiteten Variablen, bei möglichst kleiner Fehlervarianz - oder, was auf dasselbe hinausläuft, mit möglichst hoher Korrelation - zu beschreiben. Man konstruiert zu diesem Zweck jeweils Linearkombinationen aus den Elementen der beiden Variablenmengen, so daß die Korrelation zwischen den beiden Linearkombinationen maximal ist.

Üblicherweise wird bei dem Modell der kanonischen Korrelation davon ausgegangen, daß beide Variablenmengen Zufallsgrößen enthalten. Das bedeutet, daß in diesem Modell beide Variablenmengen als „gleichwertig“ angesehen werden, d.h. im statistischen Modell keine Variablenmenge a priori als Kriterium bzw. Prädiktor festgelegt ist. Die Beziehungen zwischen den beiden Variablenmengen werden also als ungerichtet bzw. symmetrisch angesehen. Wir werden auch hier diese Darstellungsform wählen. Analog zur klassischen Korrelation kann man die kanonische Korrelation aber auch so formulieren, daß nur die Variablen der einen Variablenmenge (Menge der Kriteriumsvariablen) als zufällige Größen aufgefaßt werden, während die Werte der Prädiktorvariablen fest vorgegebene Größen sind. Hier wird also im Modell bereits

festgelegt, welche Variablenmenge als Kriterium und welche als Prädiktor anzusehen ist. Der Zusammenhang zwischen den beiden Modellen läßt sich dadurch herstellen, daß das Modell mit fest vorgegebenen Prädiktorwerten als bedingte Betrachtungsweise des Modells mit zwei Mengen von zufälligen Größen angesehen werden kann. Bei Anderson (1958, p. 296-297) wird der Zusammenhang zwischen den beiden Modellen dargestellt.

Grundlegend für die kanonische Korrelation als statistisches Modell ist die Arbeit von Hotelling (1936), der schon zuvor (1935) ein Anwendungsbeispiel aus der Psychologie beschrieb. Weitere Anwendungsbeispiele der kanonischen Korrelation etwa aus dem Bereich der Pädagogik finden sich bei Barnett und Lewis (1963). Tintner (1946) und Bartlett (1948) zeigen Anwendungen der kanonischen Korrelation in der Ökonomie. Ein Überblick über den Zusammenhang zwischen kanonischer Korrelation, Faktorenanalyse und Diskriminanzanalyse findet sich bei McKeon (1966).

In neuerer Zeit (Vinograd, 1950) wurde die kanonische Korrelation auf mehr als zwei Variablenmengen verallgemeinert. Kettenring (1971) hat in seinem Artikel die verschiedenen Modelle der verallgemeinerten kanonischen Korrelation zusammenfassend dargestellt.

In dem Buch von Gnanadesikan (1977) findet sich eine auch für den Anwender dieser Verfahren lesbare, kompakte Darstellung der Arbeit von Kettenring, deren Nomenklatur wir in Kap. 2.6 übernehmen.

Weitere grundlegende Arbeiten zu dieser Verallgemeinerung sind die Artikel von Steel (1951), Horst (1965) und Carroll (1968). Bei Lohmöller & Oerter (1979) und Hanke et al. (1980) findet sich eine empirische Untersuchung aus dem Bereich der Pädagogik, in der die verallgemeinerte Korrelation als Analysetechnik verwendet wurde. Beiträge zur verallgemeinerten Korrelation sind in der Literatur auch unter dem Stichwort „multi set factor analysis“ zu finden.

## 2.2 Das Modell der kanonischen Korrelation für zwei Variablenmengen mit zufälligen Größen

Wir gehen von zwei Mengen von Zufallsgrößen aus, die wir in den beiden Zufallsvektoren

$$\mathbf{x}_1 = \begin{pmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1p} \end{pmatrix} \text{ und } \mathbf{x}_2 = \begin{pmatrix} x_{21} \\ x_{22} \\ \vdots \\ x_{2q} \end{pmatrix}$$

zusammenfassen, wobei wir annehmen, daß für die Anzahl der Komponenten in beiden Vektoren gilt

$$p \leq q.$$

Das Ziel der kanonischen Korrelation ist es, den Zusammenhang zwischen den beiden Variablenmengen durch Funktionen möglichst einfacher Bauart zu beschreiben, wobei ein Maximum an Korrelation zwischen den Variablenmengen erklärt werden soll. Man wählt dazu jeweils Linearkombinationen zwischen den Elementen der beiden Variablenmengen, also

$$(129) \quad v = \alpha' x_1, w = \beta' x_2$$

wobei die p- bzw. q-komponentigen Koeffizientenvektoren so zu bestimmen sind, daß die Korrelation zwischen den Zufallsvariablen v und w maximal ist. Die so konstruierten Zufallsgrößen sollen *kanonische Variablen* und der Korrelationskoeffizient zwischen beiden *kanonischen Korrelation* heißen.

Da der Wert der Korrelation zwischen den Variablen v und w sich nicht ändert, wenn wir etwa  $\alpha$  mit einem beliebigen Skalar a und  $\beta$  mit einem beliebigen Skalar b multiplizieren, müssen wir bei der Suche nach den optimalen  $\alpha$  und  $\beta$  an diese Koeffizientenvektoren weitere Bedingungen stellen. Diese an  $\alpha$  und  $\beta$  zusätzlich gestellten Normierungsbedingungen können wir auch als Nebenbedingungen an Parameter der Zufallsgrößen v und w formulieren. Das Kriterium zur Konstruktion der kanonischen Variablen ist also:

Bestimme die Koeffizientenvektoren  $\alpha$  und  $\beta$  der Linearkombinationen  $v = \alpha' x_1$  und  $w = \beta' x_2$  so, daß  $\rho(v, w)$  unter den Nebenbedingungen  $\text{var}(v) = \text{var}(w) = 1$  die maximale Korrelation zwischen den beiden Zufallsgrößen v und w ergibt.

Bezeichnen wir mit

$\Sigma_{11}$  die Kovarianzmatrix des Zufallsvektors  $x_1$

$\Sigma_{22}$  die Kovarianzmatrix des Zufallsvektors  $x_2$

$\Sigma_{12}$  die Kovarianzmatrix zwischen den Zufallsvektoren  $x_1$  und

$\Sigma_{21}^{x_2}$  die Kovarianzmatrix zwischen den Zufallsvektoren  $x_2$  und  $x_1$  wobei  $\Sigma_{21} = \Sigma_{12}'$  gilt,

so können wir das Maximierungsproblem auch so formulieren:

Zu maximieren ist  $\rho(v, w) = \alpha' \Sigma_{12} \beta$  unter den Nebenbedingungen  $\text{var}(v) = \alpha' \Sigma_{11} \alpha = 1$ ;  $\text{var}(w) = \beta' \Sigma_{22} \beta = 1$ .

Aus der mathematischen Ableitung ergibt sich dann das Problem, die Gleichungssysteme

$$\begin{aligned} \Sigma_{12}\beta - \varrho \Sigma_{11}\alpha &= 0 \\ \Sigma_{21}\alpha - \varrho \Sigma_{22}\beta &= 0 \\ (130) \quad \alpha' \Sigma_{11} \alpha &= 1 \\ \beta' \Sigma_{22} \beta &= 1 \end{aligned}$$

nach  $\alpha$  und  $\beta$  aufzulösen, wobei  $\varrho$  abkürzend für den Wert des Korrelationskoeffizienten an der Stelle des Maximums steht.

Nach einigen algebraischen Umformungen dieser Gleichungssysteme (eine ausführliche Darstellung findet man bei Timm, 1975, p. 348-349 oder bei Anderson, 1958, p. 288-296) erhält man

$$(131) \quad \begin{aligned} (\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} - \varrho^2 I) \alpha &= 0 \\ (\Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} - \varrho^2 I) \beta &= 0. \end{aligned}$$

Damit wurde das Auffinden der zur maximalen Korrelation gehörenden Koeffizienten  $\alpha$  und  $\beta$  auf ein Eigenwertproblem zurückgeführt. Eine zusammenfassende Darstellung des Eigenwertproblems findet man in fast allen Statistikbüchern zur multivariaten Analyse und zur linearen Algebra (z.B. Oberhofer, 1979). Man beachte, daß hier die zusätzlichen Modellannahmen eingegangen sind, daß  $\Sigma_{11}$  und  $\Sigma_{22}$  nichtsingulär sind.

Bevor wir angeben, wie aus (131) die gesuchten optimalen  $\alpha$  und  $\beta$  ermittelt werden, müssen wir einige allgemeine Bemerkungen zu diesem Eigenwertproblem machen. Für die beiden Eigenwertgleichungen ergeben sich jeweils gleich viele, identische, von Null verschiedene Eigenwerte. Die Anzahl der von Null verschiedenen Eigenwerte ist gleich dem Rang der Kovarianzmatrix  $\Sigma_{12}$  (Rao, 1973, p. 496; Anderson, 1958, p. 293-294). Nehmen wir an, daß der Rang von  $\Sigma_{12}$  gleich  $k$  ist, dann gibt es also zur ersten Eigenwertgleichung  $p - k$  und zur zweiten  $q - k$  viele Eigenwerte mit dem Wert Null.

Zur Bestimmung der gesuchten Koeffizientenvektoren  $\alpha$  und  $\beta$  benötigen wir also nur das erste Gleichungssystem in (131). Man bestimmt den größten Eigenwert  $\varrho_1^2$  (wir nehmen an, daß die Eigenwerte der Größe nach geordnet sind:  $\varrho_1^2 \geq \varrho_2^2 \geq \dots \geq \varrho_k^2$ ) die positive Wurzel aus dem maximalen Eigenwert heißt *erste kanonische Korrelation*. Dann bestimmt man den zum maximalen Eigenwert gehörenden Eigenvektor  $\alpha_1$  und berechnet schließlich aus der Beziehung

$$(132) \quad \beta_1 = \frac{1}{\varrho_1} \Sigma_{22}^{-1} \Sigma_{21} \alpha_1$$

den Eigenvektor, der zum maximalen Eigenwert im zweiten Gleichungssystem in (131) gehört.

Die beiden Linearkombinationen

$$(133) \quad v_1 = \alpha'_1 x_1, \quad w_1 = \beta'_1 x_2$$

sind dann die gesuchten Variablen mit maximaler Korrelation,

Zur praktischen Lösung des obigen Eigenwertproblems führt man die Eigenwertgleichungen (131) in Eigenwertgleichungen mit symmetrischen Matrizen über. Dies geschieht mit Hilfe der Faktorisierung von symmetrischen Matrizen nach Cholesky, wonach sich jede symmetrische Matrix als Produkt einer unteren Dreiecksmatrix und ihrer Transponierten darstellen lässt. Setzen wir also für  $\Sigma_{11} = TT'$  und  $\Sigma_{22} = FF'$  so ergeben sich mit

$$(134) \quad \gamma_1 = T' \alpha_1 \text{ bzw. } \delta_1 = F' \beta_1$$

anstelle von (131) die beiden Gleichungssysteme

$$(135) \quad \begin{aligned} (T^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}(T')^{-1} - \varrho^2 I)\gamma_1 &= 0, \\ (F^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}(F')^{-1} - \varrho^2 I)\delta_1 &= 0. \end{aligned}$$

Mit der Ableitung der beiden kanonischen Variablen ist eigentlich die gestellte Aufgabe, zwei Linearkombinationen aus den beiden Variablenmengen mit maximaler Korrelation zu konstruieren, gelöst. Nun ergibt sich in praktischen Anwendungsfällen häufig das Problem, daß durch diese beiden Linearkombinationen nicht sehr viel der gemeinsamen Korrelation zwischen den beiden Variablenmengen erklärt werden kann. Das bedeutet, daß die konstruierten kanonischen Variablen nur einen geringen Erklärungswert für das Prädiktionsproblem haben. Nun haben wir oben festgestellt, daß es insgesamt  $k$  von Null verschiedene Eigenwerte und damit  $k$  von Null verschiedene Korrelationen gibt, die den Zusammenhang zwischen den beiden Variablenmengen beschreiben. Es liegt also nahe, weitere Linearkombinationen der Variablenmengen mit Hilfe der zu diesen Eigenwerten gehörenden Eigenvektoren zu konstruieren. Die auf diese Weise konstruierten Linearkombinationen  $v_i = \alpha'_i x_1$ ,  $w_j = \beta'_j x_2$  haben die Eigenschaften

$$(I) \quad \text{cov}(v_i, v_j) = \begin{cases} 1 & \text{für } i = j \\ 0 & \text{für } i \neq j \end{cases}$$

$$(II) \quad \text{cov}(w_i, w_j) = \begin{cases} 1 & \text{für } i = j \\ 0 & \text{für } i \neq j \end{cases}$$

$$(III) \quad \text{cov}(v_i, w_j) = \begin{cases} \varrho_i \neq 0 & \text{für } i = 1, 2, \dots, k \\ 0 & \text{für } i > k \end{cases}$$

$$(IV) \quad \text{cov}(v_i, w_j) = 0 \text{ für } i \neq j$$

(Zum Beweis: siehe z.B. Rao, 1973, p. 496). Insgesamt können wir dann  $k$  solcher Linearkombinationen konstruieren, die vollständig den korrelativen

Zusammenhang zwischen den beiden Variablenmengen beschreiben. Man nennt die zum i-ten größten Eigenwert gehörenden Linearkombinationen

$$(136) \quad v_i = \alpha'_i x_1, \quad w_i = \beta'_i x_2$$

die *i-ten kanonischen Variablen*.

Zusammenfassung der Eigenschaften aller möglichen Linearkombinationen, die mit Hilfe aller Eigenvektoren  $\alpha_i$  ( $i = 1, 2, \dots, p$ ) und  $\beta_j$  ( $j = 1, 2, \dots, q$ ) gebildet werden können:

Berücksichtigt man noch die zu den Eigenwerten mit dem Wert Null gehörenden Eigenvektoren und faßt man alle Eigenvektoren  $\alpha_i$  und  $\beta_j$  zu den Matrizen

$$(137) \quad A = (\alpha_1, \alpha_2, \dots, \alpha_p), \quad B = (\beta_1, \beta_2, \dots, \beta_q)$$

zusammen<sup>2)</sup>, so ergibt sich wegen der Unkorreliertheit der Linearkombinationen für die Kovarianzmatrizen von

$$A'x_1 \text{ und } B'x_2$$

$$A'\Sigma_{11}A = I \text{ und } B'\Sigma_{22}B = I.$$

Für die Kovarianzmatrix zwischen  $A'x_1$  und  $B'x_2$  erhalten wir eine Matrix der Form

$$L = \underbrace{\begin{pmatrix} \rho_1 & 0 & \dots & 0 & 0 & \dots & 0 & \dots & 0 \\ 0 & \rho_2 & \dots & 0 & 0 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \dots & \rho_k & 0 & \dots & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & \dots & 0 \end{pmatrix}}_{q} \quad \underbrace{\left. \begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \end{array} \right\} p}_{q}$$

Die Transformation der ursprünglichen Variablenvektoren  $x_1$  und  $x_2$  mit Hilfe der oben gefundenen Eigenvektoren bewirkt also, daß die ursprüngliche Kovarianzmatrix

<sup>2)</sup> Wenn  $\Sigma_{12}$  den Rang  $k < p$  hat, dann sind die Koeffizientenmatrizen A und B nicht eindeutig bestimmt. Diese Unbestimmtheit beinhaltet, daß A und B nur bestimmt sind bis auf orthogonale Transformationen. Diese Unbestimmtheit kann durch zusätzliche Forderungen an die Matrizen A und B beseitigt werden. Dieses Problem ist analog dem *Rotationsproblem* im Modell der Faktorenanalyse.

$$(139) \quad \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \text{ transformiert wird in } \begin{pmatrix} I_p & L \\ L' & I_q \end{pmatrix}$$

## 2.3 Schätzung der kanonischen Korrelationen und der Koeffizientenvektoren der kanonischen Variablen

Wir nehmen an, daß von den Zufallsvektoren  $x_1$  und  $x_2$  jeweils N Realisierungen vorliegen. Fassen wir diese Realisierungen in den Datenmatrizen  $X_{1(Nxp)}$  und  $X_{2(Nxq)}$  zusammen, so erhalten wir für die Stichprobenmatrix bei entsprechender Aufspaltung wie in (139)

$$(140) \quad S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$$

mit

$$(141) \quad \begin{aligned} S_{11} &= \frac{1}{N} X'_1 \left( I - \frac{1}{N} J \right) X_1 \\ S_{12} = S_{21}' &= \frac{1}{N} X'_1 \left( I - \frac{1}{N} J \right) X_2 \\ S_{22} &= \frac{1}{N} X'_2 \left( I - \frac{1}{N} J \right) X_2 \end{aligned}$$

wobei J die Einsmatrix vom Typ (NxN) ist.

Unter der Annahme, daß die Zufallsvektoren  $x_1, x_2$  gemeinsam normalverteilt sind, ist S der Maximum-Likelihood-Schätzer für die Kovarianzmatrix  $\Sigma$ . Der Schätzer für  $A, B$  und für die kanonischen Korrelationen  $Q_i$  ergeben sich aus (131) bzw. (135), indem wir die Populationsparameter durch die jeweiligen Stichprobengrößen ersetzen und dieselben algebraischen Umformierungen vornehmen wie in Abschnitt 2.2. Unter Beachtung der Bemerkungen in Fußnote <sup>2)</sup> ergeben sich Schätzer für  $A, B$  und die  $Q_i$ , die zugleich Maximum-Likelihood-Schätzer für diese Parameter sind (vgl. dazu Anderson, 1958, p. 299). Die Verteilung dieser Schätzer wird bei Kshirsagar (1972, p. 261-277) dargestellt. Bezeichnen wir die Maximum-Likelihood-Schätzer für  $a_i, b_i$  und  $Q_i$  mit  $a_i, b_i$  und  $r_i$ , so gelten analog wie für die Parameter folgende Beziehungen.

$$(142) \quad \begin{aligned} S_{12}b_j &= r_j S_{11}a_j \\ S_{21}a_j &= r_j S_{22}b_j \\ a'_j S_{11}a_j &= 1 \\ b'_j S_{22}b_j &= 1 \end{aligned}$$

Berechnet man die kanonischen Variablen aus der Stichprobenkorrelationsmatrix  $R$ , wobei  $R$  analog zur Stichprobenkovarianzmatrix  $S$  aufgespalten wird in

$$(143) \quad R = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix},$$

so gelten folgende Beziehungen

$$(144) \quad \begin{aligned} R_{12}\tilde{b}_j &= r_j R_{11}\tilde{a}_j \\ R_{21}\tilde{a}_j &= r_j R_{22}\tilde{b}_j \\ \tilde{a}_j' R_{11} \tilde{a}_j &= 1 \\ \tilde{b}_j' R_{22} \tilde{b}_j &= 1 \end{aligned}$$

mit

$$(145) \quad \tilde{a}_j = [\text{diag}(S_{11})]^{\frac{1}{2}} a_j, \quad \tilde{b}_j = [\text{diag}(S_{22})]^{\frac{1}{2}} b_j.$$

## 2.4 Test zur Bestimmung der Anzahl der kanonischen Variablen

Wir hatten bereits mehrfach festgestellt, daß die maximale Anzahl der kanonischen Variablen gleich dem Rang der Kovarianzmatrix  $\Sigma_{12}$  ist. Wenn der Rang ( $\Sigma_{12}$ ) =  $k < p$  ist, lassen sich also  $k$  kanonische Variablenpaare mit von Null verschiedenen kanonischen Korrelationen konstruieren. Wenn man in praktischen Fällen den Parameter  $k$  schätzen will, ergibt sich das folgende Problem. Auch wenn Rang ( $\Sigma_{12}$ ) =  $k < p$  ist, hat die Matrix  $S_{12}$  mit Wahrscheinlichkeit 1 vollen Rang, also  $p$ . Das bedeutet, daß sich aufgrund dieses Rangkriteriums immer  $p$  Paare kanonischer Stichprobenvariablen ergeben. Bei der Entscheidung, wie viele kanonische Variablen aus den Stichprobenwerten konstruiert werden sollen, geht man nun anders vor; man prüft zunächst die Hypothese, daß zwischen den beiden Variablenmengen kein Zusammenhang besteht. Diese Hypothese besagt, daß  $\Sigma_{12} = 0$  ist. Zur Prüfung dieser Hypothese gibt es Tests von Hotelling (1951), Pillai (1955), Roy (1957, Kap. 14) und Wilks (1932). Bei Kshirsagar (1972, p. 331-334) findet man eine kurze Diskussion dieser Tests mit weiteren Literaturhinweisen insbesondere zu Untersuchungen über die Güte dieser Tests. Allerdings gibt es für die Anwendung dieser Tests kaum Entscheidungsregeln dafür, welches der verschiedenen Testkriterien vorzuziehen ist. Da für die Prüfgröße nach Wilks eine für praktische Fälle sehr gute Chi-quadrat-Approximation von Bartlett (1938) existiert, hat sich diese Prüfgröße in der empirischen Anwendung durchgesetzt:

$$(146) \quad \Lambda = \prod_{i=1}^s (1 - r_i^2); \quad s = \min(p, q).$$

Die Chi-quadrat-Approximation von Bartlett dazu ist

$$(147) \quad \chi_B^2 = - [(N - 1) - \frac{1}{2}(p + q + 1)] \log \Lambda,$$

wobei diese Prüfgröße annähernd  $\chi^2$ -verteilt ist, mit  $p \cdot q$  Freiheitsgraden. Danach wird die Hypothese

$$H_0: \Sigma_{12} = 0 \text{ zugunsten von } H_1: \Sigma_{12} \neq 0$$

immer dann verworfen, wenn der Prüfwert

$$\chi_B^2 > \chi_a^2(pq) \text{ ist,}$$

wobei  $a$  das Signifikanzniveau ist. Eine bessere Näherung als die von Bartlett an Wilks lambda-Kriterium stammt von Rao (1973, p. 472), deren Berechnung bei Gaensslen & Schubö (1976, p. 176-177) zu finden ist. Wenn die Nullhypothese  $H_0: \Sigma_{12} = 0$ , d.h. daß kein Zusammenhang zwischen den beiden Variablenmengen besteht, zu verwerfen ist, eliminiert man den Beitrag der ersten kanonischen Korrelation von  $\Lambda$  und prüft, ob die Prüfgröße

$$(148) \quad \chi_B^2 = - \left[ (N - 1) - \frac{1}{2} (p + q + 1) \right] \log \Lambda^*$$

$$\text{mit } \Lambda^* = \prod_{i=2}^s (1 - r_i^2)$$

einen signifikanten Zusammenhang zwischen den beiden Variablenmengen liefert. Die Prüfgröße ist in diesem Fall annähernd Chiquadrat-verteilt mit  $(p - 1)(q - 1)$  Freiheitsgraden. Dieses Vorgehen wird so lange fortgesetzt, bis die Nullhypothese nicht mehr zu verwerfen ist. Daraus ergibt sich dann die Anzahl der signifikanten kanonischen Korrelationen und damit die Anzahl der signifikanten kanonischen Variablen.

Zu diesem Testvorgehen ist allerdings zu bemerken, daß bei der Durchführung mehrerer Tests mit Hilfe desselben Datenmaterials die tatsächliche Irrtumswahrscheinlichkeit größer als das vorgegebene Signifikanzniveau  $a$  ist (vgl. dazu Gaensslen & Schubö, 1976, p. 112 und Haagen & Seifert, 1979, Kap. 9).

## 2.5 Extraktions- und Redundanzmaße

Zur besseren Interpretation der kanonischen Variablen ist es von Interesse, einerseits die Korrelation jeder kanonischen Variablen einer Variablenmenge mit den einzelnen Variablen dieser Variablenmenge zu kennen. Daraus lassen sich Varianzanteile berechnen, die angeben, zu wieviel Prozent die Varianz innerhalb einer Variablenmenge der  $i$ -ten kanonischen Variablen zugerechnet werden kann. Andererseits ist man an sog. Redundanzmaßen interessiert, die angeben, wieviel Prozent der Variabilität einer Variablenmenge der Variabilität der kanonischen Variablen, die aus der anderen Variablenmenge konstruiert

wurde, zugerechnet werden kann. Diese Redundanzmaße geben einen Index für das korrelative Überlappen der beiden Variablenmengen an. Je größer der Redundanzindex, desto stärker ist die Überlappung der beiden Variablenmengen. Eine ausführliche Darstellung der hier behandelten Extraktions- und Redundanzmaße einschließlich eines Anwendungsbeispiels findet man bei Gaesslen & Schubö (1976, p. 179-191).

Die *Strukturmatrix* ist diejenige Matrix, die die Korrelation zwischen den kanonischen Variablen einer Variablenmenge und den Elementen dieser Variablenmenge beschreibt. Für jede Variablenmenge lässt sich also eine zugehörige Strukturmatrix konstruieren.

Allgemein ergibt sich für den  $p$ -komponentigen Vektor mit den Korrelationen zwischen der kanonischen Variablen  $v_i$  und den Komponenten des Zufallsvektors  $x_1$

$$\rho(x_1, v_i) = \rho(x_1, \alpha'_i x_1) = [\text{diag}(\Sigma_{11})]^{-\frac{1}{2}} \Sigma_{11} \alpha_i.$$

Entsprechend gilt für die Korrelationen zwischen der kanonischen Variablen  $w_i$  mit den Komponenten des Zufallsvektors  $x_2$

$$\rho(x_2, w_i) = \rho(x_2, \beta'_i x_2) = [\text{diag}(\Sigma_{22})]^{-\frac{1}{2}} \Sigma_{22} \beta_i.$$

Die entsprechenden Korrelationen aus den Stichprobenwerten sind dann

$$r(X_1, \hat{v}_i) = [\text{diag}(S_{11})]^{-\frac{1}{2}} S_{11} \alpha_i$$

bzw.

$$r(X_2, \hat{w}_i) = [\text{diag}(S_{22})]^{-\frac{1}{2}} S_{22} \beta_i$$

wobei hier mit  $X_1$  und  $X_2$  die Beobachtungswertmatrizen bezeichnet werden.

Für standardisierte Meßwerte  $Z_{x1}$ ,  $Z_{x2}$  erhalten wir schließlich

$$(149) \quad r(Z_{x1}, \hat{v}_i) = R_{11} \tilde{\alpha}_i, \text{ mit } \tilde{\alpha}_i = [\text{diag}(S_{11})]^{\frac{1}{2}} \alpha_i$$

bzw.

$$(150) \quad r(Z_{x2}, \hat{w}_i) = R_{22} \tilde{\beta}_i, \text{ mit } \tilde{\beta}_i = [\text{diag}(S_{22})]^{\frac{1}{2}} \beta_i$$

Die beiden Strukturmatrizen  $K_1$  und  $K_2$  sind also

$$(151) \quad K_1 = (R_{11} \tilde{\alpha}_1, R_{11} \tilde{\alpha}_2, \dots, R_{11} \tilde{\alpha}_p) = R_{11} (\tilde{\alpha}_1, \tilde{\alpha}_2, \dots, \tilde{\alpha}_p) = R_{11} \tilde{A}$$

$$K_2 = (R_{22} \tilde{\beta}_1, R_{22} \tilde{\beta}_2, \dots, R_{22} \tilde{\beta}_p) = R_{22} (\tilde{\beta}_1, \tilde{\beta}_2, \dots, \tilde{\beta}_q) = R_{22} \tilde{B}.$$

Das Skalarprodukt von  $r(Z_{x1}, \hat{v}_i)$  kann interpretiert werden als Quadrat des multiplen Korrelationskoeffizienten zwischen  $\hat{v}_i$  und den Realisierungen der

standardisierten Zufallsvariablen  $x_{11}, x_{12}, \dots, x_{1p}$ . Da die Gesamtvarianz in der ersten Variablenmenge gleich p ist, kann

$$(152) \quad v_{x_1}^2 = \frac{1}{p} [(r(\mathbf{Z}_{x_1}, \hat{v}_i))' [r(\mathbf{Z}_{x_1}, \hat{v}_i)]] = \frac{\tilde{\mathbf{a}}_i' \mathbf{R}_{11}^2 \tilde{\mathbf{a}}_i}{p}$$

auch als der Varianzanteil der ersten Variablenmenge interpretiert werden, der der i-ten kanonischen Variablen dieser Variablenmenge zugerechnet werden kann.

Analog ist

$$(153) \quad v_{x_2}^2 = \frac{1}{q} [r(\mathbf{Z}_{x_2}, \hat{w}_i)' [r(\mathbf{Z}_{x_2}, \hat{w}_i)]] = \frac{\tilde{\mathbf{b}}_i' \mathbf{R}_{22}^2 \tilde{\mathbf{b}}_i}{q}$$

der Varianzanteil der zweiten Variablenmenge, der der i-ten kanonischen Variablen der zweiten Variablenmenge zugerechnet werden kann.

Von größerer Bedeutung für die Interpretation der kanonischen Variablen sind die sog. *Redundanzmaße*, die, wie bereits einleitend erwähnt, das Ausmaß der gegenseitigen Überlappung der beiden Variablenmengen beschreiben. Man berechnet dazu zunächst die Korrelationen zwischen den kanonischen Variablen  $v_i$  und den Komponenten des anderen Zufallsvektors  $x_2$ .

Dafür ergibt sich

$$Q(x_2, v_i) = Q(x_2, \alpha_i' x_1) = [\text{diag}(\Sigma_{22})]^{-\frac{1}{2}} \Sigma_{21} \alpha_i$$

bzw. wegen  $Q_i \Sigma_{22} \beta_i = \Sigma_{21} \alpha_i$

$$Q(x_2, v_i) = Q_i \cdot [\text{diag}(\Sigma_{22})]^{-\frac{1}{2}} \Sigma_{22} \beta_i$$

wobei  $Q_i = \alpha_i' \Sigma_{12} \beta_i$ ;

analog ist

$$\begin{aligned} Q(x_1, w_i) &= Q(x_1, \beta_i' x_2) = [\text{diag}(\Sigma_{11})]^{-\frac{1}{2}} \Sigma_{12} \beta_i \\ &= Q_i [\text{diag}(\Sigma_{11})]^{-\frac{1}{2}} \Sigma_{11} \alpha_i. \end{aligned}$$

Für die Korrelationen aus den Stichprobenwerten ergibt sich dann, wenn standardisierte Meßwerte zugrunde liegen,

$$(154) \quad r(\mathbf{Z}_{x_2}, \hat{v}_i) = \mathbf{R}_{21} \tilde{\mathbf{a}}_i = r_i \mathbf{R}_{22} \tilde{\mathbf{b}}_i$$

bzw.

$$(155) \quad r(\mathbf{Z}_{x_1}, \hat{w}_i) = \mathbf{R}_{12} \tilde{\mathbf{b}}_i = r_i \mathbf{R}_{11} \tilde{\mathbf{a}}_i$$

wobei  $r_i$  die i-te kanonische Korrelation der Stichprobe bezeichnet. Das Skalarprodukt von (154) kann interpretiert werden als Quadrat des multiplen

Korrelationskoeffizienten zwischen  $\hat{v}_i$  und den Realisierungen der standardisierten Zufallsvariablen  $x_{21}, \dots, x_{2q}$ . Da die Gesamtvarianz dieser Variablenmenge gleich  $q$  ist, kann

$$(156) \quad v_{x_2 \mid v_i}^2 = \frac{1}{q} [r(\mathbf{Z}_{x_2}, \hat{v}_i)]' [r(\mathbf{Z}_{x_2}, \hat{v}_i)] = \frac{1}{q} \tilde{\mathbf{a}}_i' \mathbf{R}_{12} \mathbf{R}_{21} \tilde{\mathbf{a}}_i = \frac{1}{q} r_i^2 \tilde{\mathbf{b}}_i' \mathbf{R}_{22}^2 \tilde{\mathbf{b}}_i$$

als Varianzanteil der *zweiten* Variablenmenge interpretiert werden, der der *i*-ten kanonischen Variablen der *ersten* Variablenmenge zugeordnet werden kann.

Analog erhalten wir:

$$(157) \quad \omega_{x_1 \mid w_i}^2 = \frac{1}{p} [r(\mathbf{Z}_{x_1}, \hat{w}_i)]' [r(\mathbf{Z}_{x_1}, \hat{w}_i)] = \frac{1}{p} \tilde{\mathbf{b}}_i' \mathbf{R}_{21} \mathbf{R}_{11} \tilde{\mathbf{b}}_i = \frac{1}{p} r_i^2 \tilde{\mathbf{a}}_i' \mathbf{R}_{11}^2 \tilde{\mathbf{a}}_i$$

als den Varianzanteil der ersten Variablenmenge, der der *i*-ten kanonischen Variablen der *zweiten* Variablenmenge zugerechnet werden kann.

Summiert man die Redundanzindizes (156) bzw. (157) über alle signifikanten kanonischen Variablen, so nennt man diese Summen die *totale Redundanz* der Variablenmenge  $\{x_{21}, \dots, x_{2q}\}$  bzw.  $\{x_{11}, \dots, x_{1p}\}$ .

Zwischen den oben definierten Extraktions- und Redundanzmaßen gelten folgende Zusammenhänge, die für die praktische Berechnung nützlich sind:

$$(158) \quad v_{x_2 \mid v_i}^2 = r_i^2 \omega_{x_2}^2$$

bzw.

$$(159) \quad \omega_{x_1 \mid w_i}^2 = r_i^2 v_{x_1}^2.$$

## 2.6 Verallgemeinerung der kanonischen Korrelation auf mehr als zwei Variablenmengen

Anstelle der bisherigen zwei Zufallsvektoren seien jetzt  $m$  Zufallsvektoren

$$\mathbf{x}'_j = (x_{j1}, x_{j2}, \dots, x_{jp_j}), j = 1, 2, \dots, m$$

gegeben.

Die Zufallsvektoren seien so indiziert, daß gilt

$$p_1 \leqq p_2 \leqq \dots \leqq p_m \text{ mit } p = \sum_{j=1}^m p_j.$$

Wiederum seien  $\Sigma_{jl}$  die Kovarianzmatrizen zwischen  $x_j$  und  $x_l$ .

Die Kovarianzmatrizen für  $j = 1$  sollen nichtsingulär sein. Analog zum Fall  $m = 2$  erfolgt die Konstruktion der kanonischen Variablen in  $p_1$  (kleinste Anzahl der Komponenten unter den  $m$  Zufallsvektoren) Schritten. Auf jeder Stufe gibt es allerdings statt eines Paars jetzt  $m$  kanonische Variablen. Gesucht ist also auf der Stufe  $s$  ein Vektor

$$\mathbf{z}^{(s)\prime} = (z_1^{(s)}, z_2^{(s)}, \dots, z_m^{(s)}),$$

wobei  $z_j^{(s)}$  eine Linearkombination des Zufallsvektors  $x_j$  ist, die analog der Normierungsbedingung im gewöhnlichen Fall  $m = 2$  die Varianz 1 haben soll.

Entsprechend der Vorgehensweise im Fall  $m = 2$  werden im verallgemeinerten Modell die Linearkombinationen so bestimmt, daß bestimmte Funktionen der Korrelationsmatrix  $\Phi^{(s)}$  von  $\mathbf{z}^{(s)}$  maximiert (bzw. minimiert) werden. Seien

$$(160) \quad z_j^{(1)} = \boldsymbol{\alpha}_j^{(1)\prime} \mathbf{x}_j, \quad j = 1, 2, \dots, m$$

die gesuchten Linearkombinationen auf der ersten Stufe, so sind z.B. nach dem SUMCOR-Kriterium „sum of correlations method“ (vgl. Kettenring, 1971, p. 435), die Koeffizientenvektoren  $\boldsymbol{\alpha}_j^{(1)}$  so zu bestimmen, daß

$$(161) \quad \sum_{i=1}^m \sum_{j=i+1}^m \varphi_{ij}^{(1)} = \text{Maximum}$$

unter den Normierungsbedingungen  $\boldsymbol{\alpha}_j^{(1)\prime} \Sigma_{jj} \boldsymbol{\alpha}_j^{(1)} = 1$ , wobei

$$(162) \quad \Phi^{(1)} = \begin{pmatrix} 1 & \varphi_{12}^{(1)} & \dots & \varphi_{1m}^{(1)} \\ \varphi_{12}^{(1)} & \ddots & & \cdot \\ \vdots & & \ddots & \vdots \\ \varphi_{1m}^{(1)} & \dots & \dots & 1 \end{pmatrix}$$

die Korrelationsmatrix aller Linearkombinationen auf der ersten Stufe ist. Bei Kettenring (1971) werden insgesamt fünf solcher Optimierungskriterien zur Konstruktion der kanonischen Variablen diskutiert, die alle für den Fall  $m = 2$  mit der gewöhnlichen kanonischen Korrelation übereinstimmen.

Wir wollen hier nur auf die MAXVAR („maximum variance method“) Methode, die von Horst (1961b, 1965) stammt, eingehen, da sich diese Methode auf die Lösung von Eigenwertproblemen direkt zurückführen läßt.

Fassen wir die Koeffizientenvektoren der Linearkombinationen auf der ersten Stufe zu folgender Blockdiagonalmatrix  $\mathbf{D}_{\alpha^{(1)}}$

$$\mathbf{D}_{\alpha(1)} = \begin{pmatrix} \boldsymbol{\alpha}_1^{(1)\prime} & 0' \dots & 0' \\ 0' & \boldsymbol{\alpha}_2^{(1)\prime} \dots & 0' \\ \vdots & & \vdots \\ 0' & \dots \dots & \boldsymbol{\alpha}_m^{(1)\prime} \end{pmatrix}$$

zusammen, dann gilt mit

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \dots & \boldsymbol{\Sigma}_{1m} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \dots & \vdots \\ \vdots & \vdots & \vdots \\ \boldsymbol{\Sigma}_{m1} & \boldsymbol{\Sigma}_{m2} \dots & \boldsymbol{\Sigma}_{mm} \end{pmatrix}$$

für  $\boldsymbol{\Phi}^{(1)}$

$$(163) \quad \boldsymbol{\Phi}^{(1)} = \mathbf{D}_{\alpha(1)} \boldsymbol{\Sigma} \mathbf{D}'_{\alpha(1)}.$$

Wenn wir wie im gewöhnlichen Fall die Kovarianzmatrizen  $\boldsymbol{\Sigma}_{ij}$  mit Hilfe der Cholesky-Zerlegung darstellen durch  $\boldsymbol{\Sigma}_{ij} = T_j T_j'$  und die transformierten Variablen  $y_j = T_j^{-1} x_j$  benutzen, so ergibt sich für die Kovarianzmatrix der  $y_j$

$$\boldsymbol{\Gamma} = \begin{pmatrix} \mathbf{I} & \boldsymbol{\Gamma}_{12} & \boldsymbol{\Gamma}_{1m} \\ \boldsymbol{\Gamma}_{21} & \mathbf{I} & \vdots \\ \vdots & & \vdots \\ \boldsymbol{\Gamma}_{m1} & \boldsymbol{\Gamma}_{m2} & \mathbf{I} \end{pmatrix}$$

mit  $\boldsymbol{\Gamma}_{ij} = T_i^{-1} \boldsymbol{\Sigma}_{ij} (T_j^{-1})'$ .

Für die kanonischen Variablen  $z_j^{(1)}$  gilt dann wegen  $x_j = T_j y_j$

$$(164) \quad z_j^{(1)} = \boldsymbol{\alpha}_j^{(1)\prime} T_j y_j = \boldsymbol{\beta}_j^{(1)\prime} y_j; \text{ mit } \boldsymbol{\beta}_j^{(1)\prime} = \boldsymbol{\alpha}_j^{(1)\prime} T_j.$$

Dann ergibt sich für  $\boldsymbol{\Phi}^{(1)}$

$$(165) \quad \boldsymbol{\Phi}^{(1)} = \mathbf{D}_{\beta(1)} \boldsymbol{\Gamma} \mathbf{D}'_{\beta(1)},$$

wobei  $\mathbf{D}_{\beta(1)}$  analog als Blockdiagonalmatrix wie  $\mathbf{D}_{\alpha(1)}$  definiert ist. In diesem Fall genügen die Koeffizientenvektoren den Nebenbedingungen  $\boldsymbol{\beta}_j^{(1)\prime} \boldsymbol{\beta}_j^{(1)} = 1$ . Nach dem MAXVAR-Kriterium von Horst (1961b, 1965) sind die Koeffizientenvektoren so zu bestimmen, daß die Varianz der ersten Hauptkomponente von  $z^{(1)}$  maximal ist. Dieses Optimierungskriterium ist äquivalent damit, die zum maximalen Eigenwert von  $\boldsymbol{\Phi}^{(1)}$  gehörenden Eigenvektoren zu bestimmen. Der Zusammenhang zwischen den Eigenwerten von  $\boldsymbol{\Phi}^{(1)}$  und  $\boldsymbol{\Gamma}$  ist in Lemma 2 bei Kettenring (1971) formuliert.

Bisher wurden die kanonischen Variablen für die Grundgesamtheit formuliert. Liegt eine Stichprobe vom Umfang  $N$  für jede der  $m$  Variablenmengen vor (man beachte, daß  $N > p$  sein soll), so ersetzen wir für die Ableitung der

kanonischen Variablen, die aus den Stichprobenwerten gewonnen werden, analog zum gewöhnlichen Fall, in den für die Grundgesamtheit geltenden Beziehungen die Grundgesamtheitsparameter durch die jeweiligen Stichprobenkennwerte.

$S = (S_{ij})$  entspricht dann der Kovarianzmatrix  $\Sigma$  der Variablen  $x_j$ ,

$R = (R_{ij})$  entspricht dann der Korrelationsmatrix  $\Gamma$  der transformierten Variablen  $y_j$ ,

wobei für die einzelnen Stichproben-Kovarianzmatrizen die Cholesky-Zerlegungen vorgenommen werden.

Sei

$$S_{jj} = E_j E_j'$$

dann gilt

$$(166) \quad R_{ij} = E_i^{-1} S_{ij} (E_j^{-1})'.$$

Bei der Berechnung der kanonischen Koeffizientenvektoren aus den Stichprobenwerten geht man dann folgendermaßen vor: Man berechnet die erste Hauptkomponente der Variablen  $y_j$ ; dazu bestimmt man den zum größten Eigenwert von  $R$  gehörenden Eigenvektor  $v^{(1)}$ .  $v^{(1)}$  wird dann partitioniert in  $m$  Teilvektoren  $v^{(1)} = (v_1^{(1)}, v_2^{(1)}, \dots, v_m^{(1)})$ , wobei jeder zur Länge Eins normiert wird. Daraus ergeben sich dann die Schätzer  $b_j^{(1)}$  als

$$(167) \quad b_j^{(1)} = \frac{v_j^{(1)}}{\|v_j^{(1)}\|}.$$

Wir haben hier nur die Bestimmung der  $m$  kanonischen Variablen auf der ersten Stufe aufgezeigt, da die Interpretation zusätzlicher kanonischer Variablen auf den weiteren Stufen in praktischen Anwendungsfällen meist auf Schwierigkeiten stößt. Ein auf Horst (1965) zurückgehendes Anwendungsbeispiel der verallgemeinerten kanonischen Korrelation wird auch bei Kettenring (1971) und in dem Buch von Gnanadesikan (1977, p. 74-78) diskutiert.

## Literatur

- Aigner, D. J. & Goldberger, A. S. 1977. Latent variables in socio-economic models. Amsterdam: North Holland.
- Anderson, O. D. 1976. Time series analysis and forecasting. The Box-Jenkins approach. London: Butterworths.
- Anderson, T. W. 1958. An introduction to multivariate statistical analysis. New York: Wiley.

- Bard, Y. 1974. Nonlinear parameter estimation. New York: Academic Press.
- Barnett, V. D. & Lewis, T. 1963. A study of the relation between G. C. E. and degree results. *Journal of the Royal Statistical Society, A*, 126, 187-195.
- Bartlett, M. S. 1938. Further aspects of the theory of multiple regression. *Proceedings of the Cambridge Philosophical Society*, 34, 33-40.
- Bartlett, M. S. 1948a. A note on the statistical estimation of demand and supply relations from time series. *Econometrica*, 16, 323.
- Bartlett, M. S. 1948b. Internal and external factor analysis. *British Journal of Psychology (Stat. Sec.)*, 1, 73-81.
- Beutel, P., Küffner, H. & Schubö, W. 1980. Statistik-Programm-System für die Sozialwissenschaften: SPSS 8. Stuttgart: Gustav Fischer.
- Beutel, P. & Schubö, W. 1982. Statistik-Programm-System für die Sozialwissenschaften: SPSS 9. Stuttgart: Gustav Fischer.
- Bortz, J. 1977. Lehrbuch der Statistik. Für Sozialwissenschaftler. Berlin: Springer.
- Bowden, D. C. 1970. Simultaneous confidence bands for linear regression models. *Journal of the American Statistical Association*, 65, 413-421.
- Box, G. E. P. 1974. Statistics and the environment. *Journal of the Washington Academy of Sciences*, 64 (2), 52-59.
- Box, G. E. P. & Jenkins, G. H. 1970. Time series analysis. Forecasting and control. San Francisco: Holden Day.
- Box, G. E. P., Jenkins, G. H. & Batson, D. W. 1967. Models for forecasting seasonal and non-seasonal time series. In: Harris, B. (ed.), Spectral analysis of time series. New York: Wiley.
- Box, G. E. P. & Pierce, D. A. 1970. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65, 1509-1526.
- Box, G. E. P. & Tiao, G. C. 1965. A change in level of a nonstationary time series. *Biometrika*, 52, 181-192.
- Box, G. E. P. & Tiao, G. C. 1975. Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*, 70, 70-79.
- Carroll, J. D. 1968. Generalization of canonical correlation analysis to three or more sets of variables. *Proceedings of the American Psychological Association, 76th Annual Convention*, 227-228.
- Chang, J. J. & Carroll, J. D. 1964. A general index of nonlinear correlation and its application to the program of relating physical and psychological dimensions. *American Psychologist*, 19, 540-541.
- Cochran, W. G. 1970. Some effects of errors of measurement on multiple correlation. *Journal of the American Statistical Association*, 65, p. 22-34.
- Cohen, J. & Cohen, P. 1975. Applied multiple regression/correlation analysis for the behavioral sciences. Hillsdale: Lawrence Erlbaum.

- Conger, A. J. 1974. A revised definition for suppressor variables: A guide to their identification and interpretation. *Educational and Psychological Measurement*, 34, 35-46.
- Conger, A. J. & Jackson, D. M. 1972. Suppressor variables, prediction, and the interpretation of psychological relationships. *Educational and Psychological Measurement*, 32, 579-599.
- Cooley, W. W. & Lohnes, P. R. 1971. *Multivariate data analysis*. New York: Wiley.
- Dahme, B. 1975. Zeitreihenanalyse des psychotherapeutischen Prozesses. In: Tack, W. H. (Hrsg.). *Bericht über den 29. Kongreß der Deutschen Gesellschaft für Psychologie*. Göttingen: Hogrefe.
- Dahme, B. 1976. Beziehungen zwischen Modellen der mathematischen Lern- und Entscheidungstheorien und ARIMA-Modellen. Konsequenzen für die Psychotherapie-Prozeßforschung. *Arbeitspapier Medizinische Psychologie*, Universität Hamburg.
- Darlington, R. B. 1968. Multiple regression in psychological research and practice. *Psychological Bulletin*, 69 (3), 161-182.
- Dhrymes, Ph. 1978. *Introductory Economics*. New York: Springer.
- Dixon, W. J. & Brown, M. B. 1981. *BMDP: Biomedical Computer programs P-series*. Berkeley: University of California Press.
- Draper, N. R. & Smith, H. 1966. *Applied regression analysis*. New York: Wiley.
- Durbin, J. & Watson G. 1950. Testing for serial correlation in least squares regression I. *Biometrika*, 37, 402-428.
- Durbin, J. & Watson, G. 1951. Testing for serial correlation in least squares regression II. *Biometrika*, 38, 159-178.
- Durbin, J. & Watson, G. 1971. Testing for serial correlation in least squares regression III. *Biometrika*, 58, 1-19.
- Edwards, A. L. 1979. *Multiple regression and the analysis of variance and covariance*. San Francisco: Freeman.
- Finn, J. D. 1974. *A general model for multivariate analysis*. New York: Holt, Rinehart and Winston.
- Frane, J. W. 1976. Some simple procedures for handling missing data in multivariate analysis. *Psychometrika*, 41, 409-415.
- Fuller, W. A. 1976. *Introduction to statistical time series*. New York: Wiley.
- Furnival, G. M. 1971. All possible regressions with less computations. *Technometrics*, 13, 403-408.
- Gaensslen, H. & Schubö, W. 1976. *Einfache und komplexe statistische Analyse*. München: Ernst Reinhardt.
- Garside, M. J. 1965. The best subset in multiple regression analysis. *Applied Statistics Journal of the Royal Statistical Society, Series C*, 14, 196-200.
- Gartside, P. S. 1972. A study of methods for comparing several variances. *Journal of the American Statistical Association*, 67, 342-346.

- Gigerenzer, G. 1981. Messung und Modellbildung in der Psychologie. München: Ernst Reinhardt.
- Glass, G. V., Willson, V. L. & Gottman, J. M. 1975. Design and analysis of time-series experiments. Boulder, Colorado: Colorado Associated University Press.
- Gnanadesikan, R. 1977. Methods for statistical data analysis of multivariate observations. New York: Wiley.
- Gordon, R. A. 1968. Issues in multiple regression. *American Journal of Sociology*, 73, 592-616.
- Graybill, F. A. 1976. Theory and application of the linear model. North Scituate: Duxbury.
- Gudat, U. & Revenstorff, D. 1976. Interventionseffekte in klinischen Zeitreihen. *Archiv für Psychologie*, 128, 16-44.
- Haagen, K. & Pertler, R. 1976. Methoden der Statistik. Band I. Stuttgart: Kohlhammer.
- Haagen, K. & Seifert, H. G. 1979. Methoden der Statistik für Psychologen. Band II. Stuttgart: Kohlhammer.
- Hanke, B., Lohmöller, J. B. & Mandl, H. 1980. Schülerbeurteilung in der Grundschule. München: Oldenbourg.
- Hannan, E. J. 1955. Exact tests for serial correlation. *Biometrika*, 42, 133-142.
- Hannan, E. J. 1960. Time series analysis. London: Methuen.
- Hart, B. I. & von Neumann, J. 1942. Tabulation of the probabilities for the ratio of the mean square successive difference to the variance. *Annals of Mathematical Statistics*, 13, 207-214, 445-447.
- Hays, W. L. 1963. Statistics for psychologists. New York: Holt, Rinehart & Winston.
- Hersen, M. & Barlow, D. 1976. Single case experimental designs: strategies for studying behavior change in the individual. New York: Pergamon.
- Hoadley, B. 1970. A bayesian look at inverse linear regression. *Journal of the American Statistical Association*, 65, 356-369.
- Hocking, R. R. 1972. Criteria for selection of a subset regression: Which one should be used? *Technometrics*, 14, 967-970.
- Hocking, R. R. & Leslie, R. N. 1967. Selection of the best subset in regression analysis. *Technometrics*, 9, 531-540.
- Hoerl, A. E. & Kennard, R. W. 1970. Ridge regression: biased estimation of nonorthogonal Problems. *Technometrics*, 12, 55-67.
- Holling, H. 1980. Theoretische und empirische Analysen zum Suppressorkonzept. Berlin: Phil. Dissertation.
- Holtzmann, W. 1963. Statistical models for the study of change in the single case. In: Harris, C. W. (ed.). Problems in measuring change. Madison, Wisc.: University of Wisconsin Press.
- Hope, K. 1975. Methoden multivariater Analyse. Weinheim: Beltz.

- Horst, P. 1961a. Relations among m sets of measures. *Psychometrika*, 26, 129-150.
- Horst, P. 1961b. Generalized canonical correlations and their applications to experimental data. *Journal of Clinical Psychology (Monograph Supplement)*, 14, 331-347.
- Horst, P. 1965. Factor analysis of data matrices. New York: Holt, Rinehart & Winston.
- Hotelling, H. 1935. The most predictable criterion. *Journal of Educational Psychology*, 26, 139-142.
- Hotelling, H. 1936. Relations between two sets of variates. *Biometrika*, 28, 321-377.
- Hotelling, H. 1951. A generalized T-test and measure of multivariate dispersion. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, 23-41.
- Huber, H. P. 1967. Zeitreihenanalyse im diagnostischen Einzelfall. In: Merz, F. (Hrsg.). *Bericht über den 25. Kongreß der Deutschen Gesellschaft für Psychologie*. Münster 1966. Göttingen: Hogrefe.
- Huber, H. P. 1973. Psychometrische Einzelfalldiagnostik. Weinheim: Beltz.
- Hummell, H. J. & Ziegler, R. 1976. Korrelation und Kausalität. Stuttgart: Enke.
- IMSL. 1979. The IMSL library. Fortran subroutines in the areas of mathematics and statistics. Houston: IMSL.
- Jahn, J. 1979. Zur Bewertung von Schätzungen bei der gewöhnlichen Regressionsanalyse. Darmstadt: Fachbereich Mathematik der Technischen Hochschule.
- Jenkins, G. M. & Watts, D. G. 1968. Spectral analysis and its applications. San Francisco: Holden Day.
- Jöreskog, K. G. 1973. Analyzing psychological data by structural analysis of covariance matrices. In: Goldberger, A. S. & Duncan, O. D. (ed.). *Structural analysis of covariance matrices*. New York: Seminar Press, 85-112.
- Jöreskog, K. G. 1977. Structural equation models in the social sciences: Specification, estimation and testing. In: Krishnaiah, P. R. (ed.). *Applications of statistics*. Amsterdam: North Holland.
- Jöreskog, K. G. 1978. Structural analysis of covariance and correlation matrices. *Psychometrika*, 43, 443-477.
- Johnston, J. 1963. Econometric methods. New York: McGraw Hill.
- Keeser, W. 1981. Zeitreihenanalyse in der klinischen Psychologie. Ein empirischer Beitrag zur Box-Jenkins Methodologie. Berlin: Springer.
- Kendall, M. G. 1976. Time-series. London: Griffin.
- Kendall, M. G. & Stuart, A. 1976. The advanced theory of statistics. Vol. 2. London: Griffin.
- Kendall, M. G. & Stuart, A. 1966. The advanced theory of statistics. Vol. 3. New York: Hafner.
- Kerlinger, F. N. & Pedhazur, E. J. 1973. Multiple regression in behavioral research. New York: Holt, Rinehart & Winston.

- Kettenring, J. R. 1971. Canonical analysis of several sets of variables. *Biometrika*, 58, 483-451.
- Klein, L. 1953. A textbook of econometrics. Evanston: Row, Peterson & Co.
- Kmenta, J. 1971. Elements of econometrics. New York: Macmillan.
- Koerts, J. & Abrahamse, A. P. 1969. On the theory and application of the general linear model. Rotterdam: University Press.
- Kruskal, J. B. 1971. Monotone regression: continuity and differentiability properties. *Psychometrika*, 36, 57-62.
- Krutchkoff, R. G. 1967. Classical and inverse regression - methods of calibration. *Technometrics*, 9, 425-439.
- Kshirsagar, A. M. 1972. Multivariate analysis. Statistics: Textbooks and monographs, Vol. 2. New York: Marcel Dekker.
- Küchler, M. 1979. Multivariate Analyseverfahren. Stuttgart: Teubner.
- LaMotte, L. R. & Hocking, R. R. 1970. Computational efficiency in the selection of regression variables. *Technometrics*, 12, 83-93.
- Leeuw, J. de. 1977. Correctness of Kruskal's algorithms for monotone regression with ties. *Psychometrika*, 42, 141-144.
- Lienert, G. A. 1967. Testaufbau und Testanalyse. Weinheim: Beltz.
- Lösel, F. & Wüstendorfer, W. 1974. Zum Problem unvollständiger Datenmatrizen in der empirischen Sozialforschung. *Kölner Zeitschrift für Soziologie*, 26, 342-357.
- Lohmüller, J. B. 1979. Parameterschätzung für Linearstrukturbeziehungsmodelle unter partiellen Kleinstquadratkriterien. *Forschungsbericht 79.01*, Fachbereich Pädagogik der Hochschule der Bundeswehr München, Neubiberg.
- Lohmöller, J. B. & Oerter, R. (Hrsg.). 1979. Medien in der Erziehungsausbildung. München: Oldenbourg.
- Lovell, M. C. & Prescott, E. 1970. Multiple regression with inequality constraints: pretesting bias, hypothesis testing and efficiency. *Journal of the American Statistical Association*, 65, 913-925.
- Maddala, G. S. 1977. Econometrics. New York: McGraw-Hill.
- Malinvaud, E. 1966. Statistical methods of econometrics. Amsterdam: North Holland.
- Marquardt, D. 1963. An algorithm for least-squares estimation of nonlinear parameters. *Journal of SIAM*, 11, 431-441.
- McCleary, R. & Hay, R. A. 1980. Applied time series analysis for the social sciences. London: Sage Publications.
- McDonald, R. P. 1974. Testing pattern hypotheses for covariance matrices. *Psychometrika*, 39, 189-201.
- McGee, V. E. & Carleton, W. T. 1970. Piecewise regression. *Journal of the American Statistical Association*, 65, 1109-1124.
- McKeon, J. J. 1966. Canonical analysis: Some relations between canonical correlation, factor analysis, discriminant analysis, and scaling theory. *Psychometric Monograph*, 13.

- Milliken, G. A. & Graybill, F. A. 1970. Extensions of the general linear hypothesis model. *Journal of the American Statistical Association*, 65, 797-807.
- Moosbrugger, H. 1978. Multivariate statistische Analyseverfahren. Stuttgart: Kohlhammer.
- Moosbrugger, H. 1980a. Lineare statistische Modelle zur Beschreibung linearer und nichtlinearer multipler Variablenzusammenhänge. Frankfurt/Main: Institut für Psychologie der Johann-Wolfgang-Goethe-Universität.
- Moosbrugger, H. 1980b. Zur differentiellen Prognostizierbarkeit bei nichtlinearen Test-Kriterium-Zusammenhängen. Frankfurt/Main: Institut für Psychologie der Johann-Wolfgang-Goethe-Universität.
- Morgan, J. A. & Tatar, J. F. 1972. Calculation of the residual sum of squares for all possible regressions. *Technometrics*, 14, 317-325.
- Morrison, D. F. 1976. Multivariate statistical methods. New York: McGraw-Hill.
- Namboodiri, N. K., Carter, L. F. & Blalock, H. M. 1975. Applied multivariate analysis and experimental design. New York: McGraw-Hill.
- Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K. & Bent, D. H. 1975. SPSS: Statistical package for the social sciences. New York: McGraw-Hill.
- Oberhofer, W. 1979. Lineare Algebra. München: Oldenbourg.
- Olkin, I. & Pratt, J. W. 1958. Unbiased estimation of certain correlation coefficients. *Annals of Mathematical Statistics*, 29, 201-211.
- Overall, J. E. & Klett, C. J. 1972. Applied multivariate analysis. New York: McGraw-Hill.
- Petermann, F. (Hrsg.). 1977a. Psychotherapieforschung. Weinheim: Beltz.
- Petermann, F. (Hrsg.). 1977b. Methodische Grundlagen Klinischer Psychologie. Weinheim: Beltz.
- Petermann, F. 1978. Veränderungsmessung. Stuttgart: Kohlhammer.
- Pillai, K. C. S. 1955. Some new test criteria in multivariate analysis. *Annals of Mathematical Statistics*, 26, 117-121.
- Ramsey, J. B. 1969. Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society, Series B*, 31, 350-371.
- Rao, C. R. 1973. Lineare statistische Methoden und ihre Anwendungen. Berlin: Akademie-Verlag.
- Revenstorff, D. 1979. Zeitreihenanalyse für klinische Daten. Weinheim: Beltz.
- Revenstorff, D. & Keeser, W. 1979. Zeitreihenanalyse von Therapieverläufen. In: Petermann, F. & Hehl, F. (Hrsg.). Einzelfallanalyse. München: Urban & Schwarzenberg.
- Rock, D. A., Werts, C. E., Linn, R. L. & Jöreskog, K. G. 1977. A maximum likelihood solution to the errors in variables and errors in equations model. *Multivariate Behavioral Research*, 12, 187-197.
- Roy, S. N. 1957. Some aspects of multivariate analysis. New York: Wiley.

- Schach, S. & Schäfer, Th. 1978. Regressions- und Varianzanalyse. Berlin: Springer.
- Schneeweiß, H. 1974. Ökonometrie. Würzburg: Physica-Verlag.
- Schönfeld, P. 1969. Methoden der Ökonometrie. Band I. Berlin: Franz Vahlen.
- Schönfeld, P. 1971. Methoden der Ökonometrie. Band II. Berlin: Franz Vahlen.
- Schubö, W. 1980. Statistische Verlaufsanalyse am Beispiel eines Inferenztests. In: Hentschel, U. & Smith, G. (Hrsg.). Die Wahrnehmung als Zugang zu diagnostischen Problemen. Wiesbaden: Akademische Verlagsgesellschaft.
- Schubö, W. 1982. Faktorenanalyse in der psychiatrischen Forschung. München: Phil. Dissertation.
- Schubö, W. & Hentschel, U. 1978. Improved reliability estimates for the serial color-word test. Scandinavian Journal of Psychology, 19, 91-95.
- Schuchard-Fischer, C., Backhaus, K., Humme, U., Lohrberg, W., Plinke, W. & Schreiner, W. 1980. Multivariate Analysemethoden; eine anwendungsorientierte Einführung. Berlin: Springer.
- Seibel, H. D. & Nygreen, G. T. 1972. Pfadanalyse. Ein statistisches Verfahren zur Untersuchung linearer Kausalmmodelle. Zeitschrift für Sozialpsychologie, 3, 5-12.
- Seifert, H.-G. 1975. Statistische Methoden der Ökonometrie. Unveröffentlichtes Vorlesungsmanuskript. Regensburg: Lehrstuhl für Ökonometrie.
- Sievers, W. 1977. über Dummy-Variablen-Kodierung in der Varianzanalyse. Psychologische Beiträge, 19, 3, 459-462.
- Silvey, S. D. 1969. Multicollinearity and imprecise estimation. Journal of the Royal Statistical Society, Series B, 31, 539-552.
- Steel, R. G. D. 1951. Minimum generalized variance for a set of linear functions. Annals of Mathematical Statistics, 22, 456-460.
- Steward, D. & Love, W. 1968. A general canonical correlation index. Psychological Bulletin, 70, 160-163.
- Steyer, R. 1979. Rechenregeln für bedingte Erwartungen. Frankfurt/Main: Institut für Psychologie der Johann-Wolfgang-Goethe-Universität.
- Stiefel, E. 1961. Einführung in die numerische Mathematik. Stuttgart: Teubner.
- Stroud, T. W. F. 1974. Comparing regressions when measurement error variances are known. Psychometrika, 39, 53-68.
- Sulz, K.-D. 1980. Dimensionale Analyse kognitiver Konzepte. München: Phil. Dissertation.
- Tatsuoka, M. M. 1971. Multivariate analysis: Techniques for educational and psychological research. New York: Wiley.
- Theil, H. 1971. Principles of econometrics. New York: Wiley.
- Theobald, C. M. 1974. Generalisation of mean square error applied to ridge regression. Journal of the Royal Statistical Society, Series B, 36, 103-106.
- Thorndike, R. M. 1978. Correlational procedures for research. New York: Gardner Press.

- Timm, N. H. 1970. The estimation of variance-covariance and correlation matrices from incomplete data. *Psychometrika*, 35, 417-437.
- Timm, N. H. 1975. Multivariate analysis with applications in education and psychology. Belmont, California: Wadsworth.
- Tintner, G. 1946. Some applications of multivariate analysis in economic data. *Journal of the American Statistical Association*, 41, 472.
- van de Geer, J. B. 1971. Introduction to multivariate analysis for the social sciences. San Francisco: Freeman.
- Velicer, W. F. 1978. Suppressor variables and the semipartial correlation coefficient. *Educational and Psychological Measurement*, 38, 953-958.
- Vetter, H. 1972. Dynamische und statistische Kausalanalyse. *Zeitschrift für Sozialpsychologie*, 3, 13-22.
- Vinod, H. D. 1978. Ridge regression and related techniques: A survey. *The Review of Economics and Statistics*, 60, 121-131.
- Vinograde, B. 1950. Canonical positive definite matrices under internal linear transformations. *Proceedings of the American Mathematical Society*, 1, 159-161.
- Wainer, H. 1971. Piecewise regression: a simplified procedure. *The British Journal of Mathematical and Statistical Psychology*, 24, 83-92.
- Wainer, H. & Thissen, D. 1976. Three steps towards robust regression. *Psychometrika*, 41, 9-34.
- Wilks, S. S. 1932. Certain generalizations in the analysis of variance. *Biometrika*, 24, 471-494.
- Wilson, K. V. 1973. Linear regression equations as behavior models. In: Royce, J. R. (ed.). *Multivariate analysis and psychological theory*. London: Academic Press.
- Wonnacott, R. J. & Wonnacott, Th. H. 1970. *Econometrics*. New York: Wiley.
- Young, F. W., de Leeuw, J. & Takane, Y. 1976. Regression with qualitative and quantitative variables: An alternating least squares method with optimal scaling features. *Psychometrika*, 41, 505-529.
- Yule, G. U. 1927. On a method of investigating periodicities in disturbed series. With special reference to Wölfers' sunspot numbers. *Philosophical Transactions of the Royal Society, Series A*, 226, 267-298.

## 5. Kapitel

# Diskriminanzanalyse

*Joachim Krauth*

### **1. Einführung**

#### **1.1 Problemstellung**

In Kendall (1966) bzw. in Kendall & Stuart (1966, S. 314-341) wird unterschieden zwischen Diskrimination, Klassifikation und Zerlegung. Bei einem *Diskriminationsproblem* sind danach Zufallstichproben gegeben, die aus  $k$  verschiedenen Populationen stammen. Für jede Stichprobeneinheit, im weiteren *Individuum* genannt, liegen  $p$  Meßwerte vor, die zu unterschiedlichen Variablen gehören. Gesucht ist ein Verfahren, um ein neu erhobenes Individuum aufgrund der Werte, die es für die  $p$  Variablen zeigt, der richtigen aus den  $k$  Populationen zuzuweisen.

Bei einem *Klassifikationsproblem* ist eine Stichprobe von Individuen gegeben, für die jeweils  $p$  Variablen erhoben werden. Gesucht ist ein Verfahren, um zu entscheiden, ob die Individuen zu unterschiedlichen Gruppen gehören und falls dieses zutrifft, die Gruppen zu bestimmen. Bei einem *Zerlegungsproblem* schließlich ist eine Stichprobe von Individuen aus einer Population gegeben, wobei für jedes Individuum die Meßwerte für  $p$  Variablen vorliegen. Gesucht ist ein Verfahren, um die Population in Anteile mit vorgegebenen Eigenschaften zu zerlegen.

Beispielsweise möge je eine Stichprobe von Normalen, chronisch Schizophrenen und akut Schizophrenen vorliegen, d.h. es handle sich um  $k = 3$  Stichproben, wobei für jedes Individuum die Antworten auf  $p = 10$  Testitems bekannt sind. Ein neues Individuum soll aufgrund seiner Beantwortung der Testitems einer der drei Populationen zugeordnet werden. Dieses ist ein Diskriminationsproblem. Hat man jedoch eine Stichprobe von als schizophren diagnostizierten Patienten vorliegen, für die jeweils die Antworten auf zehn Testitems vorliegen, und möchte man wissen, ob es Teilpopulationen der Schizophrenen gibt, d.h. unterschiedliche Diagnosegruppen, und diese identifizieren, so liegt ein Klassifikationsproblem vor. Schließlich liege eine Stichprobe von unter

dem Verdacht einer Schizophrenie eingelieferten Personen vor und für jede Person die Antworten auf zehn Items. Falls dann die Stichprobe und damit die Population aufgrund der Itembeantwortung in eine Gruppe, die keiner Behandlung, in eine Gruppe, die einer ambulanten Behandlung, und eine Gruppe, die einer stationären Behandlung bedarf, aufgeteilt werden soll, so handelt es sich um ein Zerlegungsproblem.

Eine erste Ursache für Diskriminationsprobleme sieht Kendall (1966) in *verlorengegangenen Informationen*. Ein Beispiel dafür könnte ein psychiatrischer Patient sein, dessen Krankengeschichte abhanden gekommen ist, der aber einer Diagnosegruppe zugeteilt werden soll. Als zweite Ursache wird das Problem einer möglichst einfachen *Diagnose* angesehen, ohne daß eine langwierige Untersuchung durchgeführt werden muß. Als dritte Ursache wird die *Vorhersage* genannt, wenn z.B. aufgrund von vorliegenden Symptomen eine bestimmte Krankheitsentwicklung vorhergesagt werden soll.

Die strenge Trennung zwischen Diskrimination und Klassifikation wird nur von wenigen Autoren eingehalten. Oft werden die verschiedenen Probleme im Zusammenhang betrachtet wie z.B. bei Janke (1971). Gelegentlich wird die Diskrimination auch als Spezialfall der Klassifikation angesehen. So unterscheidet Das Gupta (1973) mehrere Probleme der Klassifikation. Das erste Problem betrifft die Zuordnung einer Zufallsstichprobe von Individuen aus der gleichen Population zu einer von  $k$  Populationen, wobei für jedes Individuum ein  $p$ -dimensionaler Meßwertvektor gegeben ist. Sind die Verteilungen zu den  $k$  Populationen nicht genau bekannt, so benötigt man sogenannte *Trainingsstichproben* aus diesen Populationen. Dieses Problem entspricht dem Diskriminationsproblem bei Kendall (1966).

Als zweites Klassifikationsproblem wird die Aufteilung einer Zufallsstichprobe auf  $k$  Populationen angesehen. Falls man die Wahrscheinlichkeiten für eine Zugehörigkeit zu den einzelnen Populationen kennt, so spricht man von *bekannter Mischverteilung*, sonst von *unbekannter Mischverteilung*. Bei unbekannter Mischverteilung und unbekannten Populationsverteilungen benötigt man Trainingsstichproben. Falls die Populationsverteilung bekannt ist, so heißt die Trainingsstichprobe *identifiziert* oder *erhärtet*. Falls die Populationsverteilung nicht bekannt ist, so heißt die Trainingsstichprobe *nichtidentifiziert*. Gelegentlich bilden die Individuen der Trainingsstichprobe eine Folge und die richtige Population kann ab einem bestimmten Folgeglied identifiziert werden. Dann heißt die Trainingsstichprobe *post-identifiziert*.

Kurze Einführungen in die Begriffsbildungen der Diskriminanzanalyse geben Trampisch (1977) und Kossack (1963). Von Alexakos (1966) und Power, Muntz & MacRae (1975) wird untersucht, wie sich Lösungen von Diskriminationsproblemen mit mathematischen Methoden von der Beurteilung durch Gruppen von klinischen Psychologen unterscheiden. Es zeigt sich, daß die

Güte der Zuordnung durchaus vergleichbar ist, wobei eine Kombination der Methoden zu einer erheblichen Verbesserung der Resultate führen kann.

## 1.2 Entstehungsgeschichte

Die ersten Vorläufer der Diskriminanzanalyse findet man in den dreißiger Jahren dieses Jahrhunderts. Die ersten Arbeiten, die das Diskriminationsproblem in der heutigen Form betreffen, sind wohl von Fisher (1936, 1938), Smith (1936-1937) und Welch (1939). Einzelheiten über diese Ansätze und über die weitere Entwicklung geben Bauer (1954), Kossack (1963) und Das Gupta (1973).

## 1.3 Übersichtsarbeiten

Der gesamte Bereich der Diskriminanzanalyse wird dargestellt in der Monographie von Lachenbruch (1975), die auch eine Bibliographie mit 579 Titeln enthält. Speziell die Diskriminanzanalyse für qualitative Daten, die sogenannte *diskrete Diskriminanzanalyse*, behandelt die Monographie von Goldstein & Dillen (1978). Die Monographie von Skarabis (1970) betrachtet in erster Linie mathematische Probleme der Diskriminanzanalyse, während der Sammelband von Cacoullos (1973 a) die unterschiedlichsten Aspekte anspricht.

Neben der Bibliographie in Lachenbruch (1975) sind diejenigen von Das Gupta (1973) und Cacoullos & Styan (1973) zu nennen. Speziell für das Problem, die Wahrscheinlichkeit für eine Fehlklassifikation zu schätzen, gibt Toussaint (1974) eine Liste von 188 Titeln an.

Übersichtsartikel über die Diskriminanzanalyse in englischer Sprache gibt es von Brown (1947), Tatsuoka & Tiedeman (1954), Isaacson (1964), Huberty (1975), Geisser (1977), Schaafsma & Vark (1977, 1979) und Lachenbruch & Goldstein (1979). In deutscher Sprache gibt Janke (1971) eine Übersicht über Klassifikationsverfahren einschließlich der Diskriminanzanalyse, während Trampisch (1975) einen Überblick über Verfahren mit nichtnormalverteilten Gruppen gibt.

## 2. Grundlagen

### 2.1 Lineare Diskriminanzfunktion

Es mögen zwei Populationen  $P_1$  und  $P_2$  mit den  $p$ -dimensionalen Verteilungen  $V_1$  und  $V_2$  für die Meßwertvektoren vorliegen. Für jedes Individuum liege ein

p-dimensionaler Meßwertvektor  $x' = (x_1, \dots, x_p)$  vor, wobei der Transpositionsstrich die Umwandlung des Spaltenvektors  $x$  in den Zeilenvektor  $x'$  bezeichnet. Gesucht ist eine *Zuordnungsregel*, die den Vektor  $x$  entweder  $P_1$  oder  $P_2$  zuordnet. Von Fisher (1936) wird vorgeschlagen, eine Linearkombination

$$y = a_1 x_1 + \dots + a_p x_p$$

zu bilden, wobei  $a_1, \dots, a_p$  so gewählt sein sollen, daß das Quadrat der Differenz der Erwartungswerte von  $Y$  in den beiden Populationen bezogen auf die gemeinsame Varianz maximal wird.

Es seien  $\mu'_1 = (\mu_{11}, \dots, \mu_{1p})$  bzw.  $\mu'_2 = (\mu_{21}, \dots, \mu_{2p})$  die Erwartungswertvektoren zu den Verteilungen  $V_1$  und  $V_2$  und  $\Sigma$  die für beide Verteilungen gleiche Kovarianzmatrix. Ohne weitere Annahmen über  $V_1$  und  $V_2$  führt das Verfahren von Fisher (1936) zu der Linearkombination

$$y = (\mu_1 - \mu_2)' \Sigma^{-1} x.$$

Falls  $\mu_1, \mu_2$  und  $\Sigma$  nicht bekannt sind, so müssen Zufallsstichproben der Größe  $n_1$  bzw.  $n_2$  aus  $P_1$  bzw.  $P_2$  vorliegen, und man schätzt  $\mu'_1$  bzw.  $\mu'_2$  durch die Mittelwertvektoren  $\bar{x}'_1 = (\bar{x}_{11}, \dots, \bar{x}_{1p})$  bzw.  $\bar{x}'_2 = (\bar{x}_{21}, \dots, \bar{x}_{2p})$  und  $\Sigma$  durch die empirische Kovarianzmatrix  $S$  aus der vereinigten Stichprobe. In diesem Fall ergibt sich

$$y = (\bar{x}_1 - \bar{x}_2)' S^{-1} x.$$

Rechenbeispiele und numerische Vereinfachungen zum Verfahren von Fisher findet man bei Beall (1945), Schmid (1950) und Behrens (1959).

Man ordnet das Individuum mit dem Wert  $y$  der Population  $P_1$  zu, falls  $y$  näher zu

$$\bar{y}_1 = (\mu_1 - \mu_2)' \Sigma^{-1} \mu_1 \text{ bzw. } \bar{y}_1 = (\bar{x}_1 - \bar{x}_2)' S^{-1} \bar{x}_1$$

liegt als zu

$$\bar{y}_2 = (\mu_1 - \mu_2)' \Sigma^{-1} \mu_2 \text{ bzw. } \bar{y}_2 = (\bar{x}_1 - \bar{x}_2)' S^{-1} \bar{x}_2.$$

Sonst ordnet man es der Population  $P_2$  zu. Dieses ist gleichbedeutend damit, daß man das Individuum der Population  $P_1$  zuordnet, falls

$$y > (\bar{y}_1 + \bar{y}_2)/2$$

gilt. Die Differenz

$$\bar{y}_1 - \bar{y}_2 = (\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 - \mu_2) = D^2 \text{ bzw.}$$

$$\bar{y}_1 - \bar{y}_2 = (\bar{x}_1 - \bar{x}_2)' S^{-1} (\bar{x}_1 - \bar{x}_2) = D_S^2$$

wurde von Mahalanobis im Jahre 1927 als Abstandsmaß zwischen zwei mehrdimensionalen Verteilungen eingeführt und stellt eine Modifikation eines Ko-

effizienten von Karl Pearson aus dem Jahre 1921 dar (vgl. Das Gupta, 1973). Mit Hilfe der bei Vorliegen von multivariaten Normalverteilungen F-verteilten Statistik

$$F = \frac{n_1 n_2 (n_1 + n_2 - p - 1)}{(n_1 + n_2)(n_1 + n_2 - 2)p} D_S^2$$

mit  $p$  und  $(n_1 + n_2 - p - 1)$  Freiheitsgraden kann man testen, ob die beiden Stichproben aus unterschiedlichen Populationen stammen.

Anders als Fisher (1936) sucht Welch (1939) eine Diskriminanzfunktion, bei der die totale *Fehlklassifikationswahrscheinlichkeit* minimal ist. Dazu seien  $q_1 = P(P_1)$  bzw.  $q_2 = P(P_2)$  die *A-priori-Wahrscheinlichkeiten*, d.h. die Wahrscheinlichkeiten dafür, daß ein Individuum zu  $P_1$  bzw.  $P_2$  gehört. Ferner sei  $R_1$  bzw.  $R_2$  der Bereich im  $p$ -dimensionalen Raum, für den Vektoren  $x$  zu  $P_1$  bzw.  $P_2$  zugeordnet werden. Dabei seien  $R_1$  und  $R_2$  elementfremde Mengen, die vereinigt den ganzen Raum ergeben. Die totale Fehlklassifikationswahrscheinlichkeit ergibt sich zu

$$q_1 P(R_2 | P_1) + q_2 P(R_1 | P_2).$$

Falls  $f_1(x)$  bzw.  $f_2(x)$  die zu  $V_1$  bzw.  $V_2$  gehörige  $p$ -dimensionale Wahrscheinlichkeitsdichte ist, so lautet die Klassifikationsregel: Ordne  $x$  der Population  $P_1$  zu, falls

$$f_1(x) / f_2(x) > q_2/q_1$$

und der Population  $P_2$  im umgekehrten Fall. Speziell für multivariate Normalverteilungen mit gleicher Kovarianzmatrix ergibt sich hieraus als optimale Zuordnungsregel: Ordne  $x$  der Population  $P_1$  zu für

$$W_T = [x - 0.5(\mu_1 + \mu_2)]' \Sigma^{-1} (\mu_1 - \mu_2) > \ln[q_2/q_1].$$

Bei unbekannten Parametern schätzt man  $W_T$  aus Stichproben und verwendet die Diskriminanzfunktion

$$W = [x - 0.5(\bar{x}_1 + \bar{x}_2)]' S^{-1} (\bar{x}_1 - \bar{x}_2).$$

Die Koeffizienten dieser linearen Diskriminanzfunktion sind identisch mit denen der Funktion von Fisher.

Eine andere Idee zur Konstruktion einer Diskriminanzfunktion stammt von Mises (1945), der vorschlägt, die maximale Fehlklassifikationswahrscheinlichkeit  $P(R_2 | P_1)$  bzw.  $P(R_1 | P_2)$  in den beiden Gruppen zu minimieren. Andere Autoren nehmen unterschiedliche Kosten für verschiedene Arten der Fehlklassifikation an und minimieren die Gesamtkosten (vgl. Anderson, 1958).

Neben der auf Anderson (1958) zurückgehenden Diskriminanzfunktion W werden in der Literatur auch die Funktionen

$$U = \mathbf{x}' S^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$$

von Wald (1944) und

$$Z = \frac{n_1}{n_1+1} (\mathbf{x} - \bar{\mathbf{x}}_1)' S^{-1} (\mathbf{x} - \bar{\mathbf{x}}_1) - \frac{n_2}{n_2+1} (\mathbf{x} - \bar{\mathbf{x}}_2)' S^{-1} (\mathbf{x} - \bar{\mathbf{x}}_2)$$

von John (1960a) betrachtet. Ein numerisches Beispiel zur Klassifikation mit Hilfe von U gibt Kossack (1945).

Während bei bekannten Parametern die obigen Funktionen normalverteilt sind für multivariat normalverteiltes X, gilt dieses, falls man für die Parameter deren Schätzungen einsetzt, nur noch für die bedingten Verteilungen bei gegebenen  $\bar{\mathbf{x}}_1$ ,  $\bar{\mathbf{x}}_2$  und S. Die nichtbedingte Verteilung ist nicht normal. Untersuchungen über das finite und asymptotische Verhalten der Verteilung von W wurden u.a. durchgeführt von Anderson (1951, 1973a, 1973b), Harter (1951), Sitgreaves (1952, 1961, 1973), Bowker & Sitgreaves (1961), Teichroew & Sitgreaves (1961), Bowker (1961), Elfving (1961), John (1962-1963), Okamoto (1963) und Lachenbruch (1968). Die Verteilung von U untersuchten u. a. Wald (1944), Sitgreaves (1952) und John (1959, 1960a, 1960b, 1962-1963, 1964), wobei der letztere allerdings ebenso wie Memon & Okamoto (1971) genaugenommen die Verteilung von Z betrachtete.

Eine Reihe von Autoren weisen Optimalitätseigenschaften der linearen Diskriminanzfunktion nach. So zeigt Kudo (1959), daß diese Funktion bei univariaten Normalverteilungen mit gemeinsamer bekannter Varianz ein Zweient-scheidungsproblem optimiert. Von Das Gupta (1965) wird die Maximum-Likelihood-Schätzung der Diskriminanzfunktion untersucht. Dazu wählt man zu vorgegebenem x die Verteilungsparameter so, daß die Wahrscheinlichkeit dafür, daß x aus einer bestimmten Population stammt, maximal wird. Man ordnet x der Population zu, für die diese Wahrscheinlichkeit maximal ist. Die Maximum-Likelihood-Regel besitzt gewisse Güteeigenschaften. Von Porebski (1966 a) wird gezeigt, daß bei gleichen A-priori-Wahrscheinlichkeiten  $q_1$  und  $q_2$  und multivariaten Normalverteilungen mit bekannter gemeinsamer Kovarianzmatrix die Fisher'sche Diskriminanzfunktion gleich einer Bayes-Lösung und gleich der Maximum-Likelihood-Lösung ist. Für zwei multivariate Normalverteilungen, die bis auf die gemeinsame Kovarianzmatrix bekannt sind, betrachtet Streit (1977) das Maß  $D^2$  und zeigt, daß die Zuordnungsregel eine Maximum-Likelihood-Regel und eine zulässige Bayes-Regel ist.

Von Matusita (1973) wird eine lineare Diskriminanzfunktion durch Minimierung eines Informationsmaßes erhalten, während Enis & Geisser (1971, 1974) unter allen linearen Diskriminanzfunktionen, die auf Parameterschätzungen

beruhen, diejenige bestimmen, die die totale Wahrscheinlichkeit der Fehlklassifikation minimiert. Für den Fall univariater Normalverteilungen mit gemeinsamer Varianz bestimmen Ganesalingam & McLachlan (1978) bei Schätzung der Parameter die optimale Diskriminanzfunktion. Ebenso wie O'Neill (1978) berechnen sie die asymptotische relative Effizienz einer linearen Diskriminanzfunktion für den Fall einer nichtklassifizierten Ausgangsstichprobe, die eine Mischung von zwei multivariaten Normalverteilungen mit gemeinsamer Kovarianzmatrix darstellt. Die Simulationsstudien von Ganesalingam & McLachlan (1979) zeigen, daß die in dieser Situation geschätzte lineare Diskriminanzfunktion auch bei kleinen Stichprobenumfängen die Populationen gut trennt, obwohl die Schätzung der Koeffizienten der Diskriminanzfunktion nicht sehr reliabel ist. Von Day (1969) wird gezeigt, daß die lineare Diskriminanzfunktion optimal ist für alle multivariaten Wahrscheinlichkeitsdichten, die zu der Klasse der linearen Exponentialverteilungen gehören.

Verschiedene Spezifikationen und auch Erweiterungen des ursprünglichen Ansatzes werden ebenfalls untersucht. So betrachtet Han (1968) den Fall gleicher Korrelationen der Variablen, Han (1969) den Fall proportionaler Kovarianzmatrizen, d.h.  $\Sigma_1 = c\Sigma_2$ , und Han (1970) den Fall zirkulärer Kovarianzmatrizen. Von Rao (1966a) sowie Lachenbruch & Kupper (1973) wird angenommen, daß  $V_1$  und  $V_2$  selber Mischungen von multivariaten Normalverteilungen sind. In McLachlan (1975b, 1977a) wird zusätzlich die Information ausgenutzt, daß außer den Meßwertvektoren der eindeutig zu  $P_1$  bzw.  $P_2$  gehörenden Individuen auch noch die Meßwertvektoren nichtklassifizierter Individuen vorliegen.

Von Burnaby (1966) werden Diskriminanzfunktionen entwickelt, die invariant sind gegen bekannte nichtdiskriminierende Variation der Variablen, wie sie z.B. bei Wachstumskurven auftritt. Cochran (1964a) zeigt, daß die Fehlklassifikationswahrscheinlichkeiten aus der Diskriminationsfähigkeit der Einzelvariablen nur schwer vorhersagbar sind und erheblich von den Interkorrelationen der Variablen abhängen. Dabei ist es von Vorteil, wenn Variablen mit guten Diskriminatoren negativ korreliert sind.

Efron (1975) betrachtet zwei multivariat normalverteilte Populationen mit einer Trainingsstichprobe  $x_1, \dots, x_n$ . Die Zuordnung eines Meßwertvektors aufgrund der Schätzung der linearen Diskriminanzfunktion von Fisher wird als normales Diskriminanzverfahren bezeichnet. Bei Schätzung der Funktion von Fisher bedingt bei gegebenen Werten von  $x_1, \dots, x_n$  wird von *logistischer Regression* gesprochen. Unter Berücksichtigung der asymptotischen Fehlerraten hat die logistische Regression asymptotische relative Effizienzen bezogen auf das normale Diskriminanzverfahren, die zwischen 50% und 66% liegen. Borgen & Seling (1978) verwenden die Diskriminanzanalyse in Anschluß an eine multivariate Varianzanalyse, um die verschiedenen Populationen zu trennen.

Die lineare Diskriminanzfunktion von Fisher (1936) wird von Travers (1939) zur Trennung von Ingenieuren und Piloten, von Jackson (1950) zur Aufteilung von Studenten auf zwei Kurse, von Ahmann (1955) zur Klassifikation von Studenten nach der Studienrichtung, von Pichot & Perse (1952) zur diagnostischen Trennung von Arterioskleroseerkrankten und Senilen sowie von Wheeler (1963) zur Trennung von Populationen Hirngeschädigter verwendet.

## 2.2 Bayes-Ansatz

Bezeichnet man mit  $f_1(x)$  bzw.  $f_2(x)$  die multivariaten Dichten zu  $V_1$  bzw.  $V_2$  und mit  $q_1$  bzw.  $q_2$  die A-priori-Wahrscheinlichkeiten, so ergeben sich nach der Formel von Bayes die *A-posteriori-Wahrscheinlichkeiten* für ein Vorliegen von  $P_i$  bei gegebenem Meßwertvektor  $x$  zu

$$P(P_i | x) = q_i f_i(x) / [q_1 f_1(x) + q_2 f_2(x)], \quad i=1,2.$$

Ordnet man  $x$  der Population  $P_1$  zu, falls

$$P(P_1 | x) > P(P_2 | x)$$

gilt, so ist diese Regel äquivalent mit der Regel, die die totale Fehlklassifikationswahrscheinlichkeit minimiert. Die A-posteriori-Wahrscheinlichkeiten sind nützlich, wenn man das Risiko eines Individuums schätzen will, zu einer bestimmten Population, z.B. einer Diagnosegruppe, zu gehören. In Ben-Bassat & Gal (1977) wird das Verhalten der A-posteriori-Wahrscheinlichkeiten bei nicht identisch verteilten Beobachtungen untersucht.

Man unterscheidet einen Semi-Bayes-Ansatz und einen vollständigen Bayes-Ansatz. Beim *Semi-Bayes-Ansatz* benutzt man A-priori-Dichten der Form  $g(\mu_1, \mu_2, \Sigma)$  für die unbekannten Parameter, um damit die A-posteriori-Verteilung der Diskriminanzfunktion bei gegebenen  $\bar{x}_1$ ,  $\bar{x}_2$  und  $S$  zu untersuchen. Beim *vollständigen Bayes-Ansatz* betrachtet man die *Vorhersage-Dichten*  $f(x | \bar{x}_1, \bar{x}_2, S, P_i)$  und verwendet als Diskriminanzfunktion

$$\ln[f(x | \bar{x}_1, \bar{x}_2, S, P_1) / f(x | \bar{x}_1, \bar{x}_2, S, P_2)].$$

Mit Hilfe dieser Funktion werden Individuen den zugehörigen Populationen zugeordnet. Eine Beschreibung des vollständigen Bayes-Ansatzes gibt Geisser (1966), während Desu & Geisser (1973) beide Ansätze für verschiedene Situationen diskutieren. Der kritische Punkt bei allen Bayes-Verfahren ist immer die Frage nach der richtigen Wahl der A-priori-Verteilungen. In Hermans & Habbema (1975) werden deshalb anhand von zwei Sätzen experimenteller Daten fünf verschiedene Methoden, die Dichten  $f_1(x)$  und  $f_2(x)$  zu schätzen, miteinander verglichen. In Geisser (1964) werden unter der Annahme, daß für  $k$  Populationen multivariate Normalverteilungen vorliegen, acht verschiedene

Konstellationen für die Parameter  $\mu_i$  und  $\Sigma_i$  betrachtet mit verschiedenen A-priori-Dichten für diese Parameter und zusätzlich fünf Situationen für die Wahrscheinlichkeiten  $q_i$ . Von Geisser & Desu (1968) wird für zwei Populationen  $P_1$  und  $P_2$ , für die  $V_1$  und  $V_2$  multivariat normal sind, der Fall  $\mu_1 = \mu_2 = 0$  betrachtet, wobei  $\Sigma_1$  und  $\Sigma_2$  jeweils konstante Varianzen und Kovarianzen enthalten sollen. Für verschiedene Wahlen dieser Varianzen und Kovarianzen und der A-priori-Dichten werden Vorhersagedichten bestimmt. In Gupta (1977) wird der Fall von multivariat normalen  $V_1, \dots, V_k$  mit unbekannten  $\mu_i$  und  $\Sigma_i$  betrachtet und gezeigt, daß zwei Zuordnungsregeln äquivalent sind, von denen die eine auf einer Diskriminanzfunktion und die andere auf einem Bayes-Ansatz mit gleichen A-priori-Wahrscheinlichkeiten  $q_i$  beruht. Hora (1978) betrachtet für zwei Populationen einmal den Fall, daß  $\mu_1$  und  $\mu_2$  unbekannt,  $\Sigma_1 = \Sigma_2$  aber bekannt ist, und nimmt für  $\mu_1$  und  $\mu_2$  normale A-priori-Dichten an. Im zweiten Fall setzt er  $\mu_1 = \mu_2$  als unbekannt mit normaler A-priori-Dichte und  $\Sigma_1 = \Sigma_2$  als unbekannt mit einer Wishart A-priori-Dichte voraus. In beiden Fällen wird der optimale Stichprobenumfang bestimmt (zum Bayes-Ansatz vgl. Kapitel 4 des Bandes „Hypothesenprüfung“ aus der Enzyklopädie der Psychologie).

## 2.3 Fehlerraten

Der Begriff der Fehlerrate wird in der älteren Literatur zur Diskriminanzanalyse durchaus in unterschiedlichem Sinne verwendet. Erst in den Arbeiten von Hills (1966), Skarabis (1970), Sorum (1972), Michaelis (1973 b), Trampisch (1975, 1977), Victor (1976a) und Lachenbruch (1975) werden die verschiedenen Fehlerarten exakt definiert und voneinander abgegrenzt.

Zunächst werde das Diskriminationsproblem für  $k$  Populationen mit vollständiger Information betrachtet. Dann sind sowohl die A-priori-Wahrscheinlichkeiten  $q_1, \dots, q_k$  als auch die zu den Verteilungen  $V_1, \dots, V_k$  gehörigen Dichten  $f_1(x), \dots, f_k(x)$  bekannt. Als tatsächlichen **Fehler**  $F(D)$  der Zuordnungsregel  $D$  bezeichnet man den Anteil der insgesamt falsch klassifizierten Fälle

$$F(D) = \sum_{i \neq j} \sum q_i P(R_j | P_i).$$

Dabei ist  $R_j$  der Bereich, für den ein Meßwertvektor  $x$  aus  $R_j$  der Population  $P_j$  zugeordnet wird. Eine Zuordnungsregel  $D_o$  heißt *optimal*, falls  $F(D_o)$  unter allen möglichen Zuordnungsregeln  $D$  minimal ist.  $F(D_o)$  heißt in diesem Fall (*theoretisch*) optimaler *Fehler*. Es gibt also keine Regel  $D$ , die insgesamt weniger Fehlklassifikationen aufweist als  $D_o$ . Jedoch muß  $D_o$  nicht eindeutig bestimmt sein. Die optimale Zuordnungsregel ist dadurch gegeben, daß ein Vektor  $x$  der Population  $P_i$  zugeordnet wird, falls  $q_i f_i(x)$  maximal ist unter allen  $q_1 f_1(x), \dots, q_k f_k(x)$ .

In der Praxis sind weder die  $q_i$  noch die  $f(x)$  bekannt, sondern man hat nur eine sogenannte *erhärtete Stichprobe* vorliegen, bei der für jedes  $x$  bekannt ist, zu welcher Population es gehört. Da die Zuordnungsregel  $D_S$  sich durch Schätzung aus der erhärteten Stichprobe ergibt, ist  $D_S$  eine Zufallsvariable, und auch der erhaltene Fehler  $F(D_S)$  der insgesamt falsch klassifizierten Fälle ist eine Zufallsvariable. Man bezeichnet den Erwartungswert  $E[F(D_S)]$  als *mittleren Fehler* von  $D_S$ .

Ist  $D_S$  eine geschätzte Zuordnungsregel und sind  $f_{is}(x)$  geschätzte Dichten, so heißt

$$F_S(D_S) = \sum_{i \neq j} \sum_{q_i} P_S(R_j | P_i)$$

*unterschätzter Fehler* oder *scheinbarer Fehler* von  $D_S$ . Dieses ist also der Anteil der ursprünglichen Stichprobe, der durch die geschätzte Diskriminanzfunktion fehlklassifiziert wird. Unter ziemlich allgemeinen Bedingungen (vgl. Glick, 1973a) gilt

$$F_S(D_S) \leq F(D_S) \text{ und } E[F_S(D_S)] \leq E[F(D_S)].$$

Sind die Dichten  $f_i$  auf bestimmte Verteilungsklassen, z.B. Normalverteilungen, eingeschränkt, in denen bestimmte Parameter nicht festgelegt sind, so bezeichnet man  $F_s(D)$  als den *erreichbaren Fehler*, falls  $F_s(D)$  unter allen zulässigen Wahlen der noch unbestimmten Parameter minimal ist. Müssen die Parameter geschätzt werden, so heißt  $F_M(D_S)$  der *tatsächlich erreichbare Fehler*.

In John (1961) werden für zwei Populationen für geschätzte Diskriminanzfunktionen bei Vorliegen multivariater Normalverteilungen die Verteilungen der Fehlerraten untersucht und die Erwartungswerte bestimmt, während in John (1964) die Zuordnungswahrscheinlichkeiten betrachtet werden für den Fall, daß  $x$  weder zu  $P_1$  noch zu  $P_2$  gehört. Von Dunn & Varady (1966) wird durch Simulationsstudien der Zusammenhang zwischen der tatsächlichen und der geschätzten Fehlerrate bestimmt. Für den Fall zweier Populationen bei Vorliegen multivariater Normalverteilungen und Verwendung der Diskriminanzfunktion  $W$  untersucht McLachlan (1972, 1974) die asymptotischen Verteilungen der *bedingten Fehlerraten*  $p_1 = P(R_2 | P_1)$  und  $p_2 = P(R_1 | P_2)$  und des bedingten Risikos  $(p_1 + p_2)/2$ , während für den Fall proportionaler Kovarianzmatrizen und geschätzter quadratischer Diskriminanzfunktion in McLachlan (1975 a) die erwarteten Fehlerraten bestimmt werden. Für das allgemeine Klassifikationsproblem, wo die Indexvariable, die in unseren bisherigen Fragestellungen die Populationen kennzeichnet, nicht mehr diskret ist, konstruieren Marshall & Olkin (1968) Verfahren, die das Risiko minimieren. Von Moran (1975) werden exakte Ausdrücke für die Erwartungswerte des tatsächlichen, des scheinbaren und des *geschätzten Fehlers*  $F_s(D)$  für die lineare Diskrimi-

nanzfunktion angegeben für den Fall, daß  $V_1$  und  $V_2$  multivariat normal sind mit gleicher Kovarianzmatrix.

Bei Benutzung der geschätzten anstelle der theoretischen Diskriminanzfunktion wird die Fehlklassifikationswahrscheinlichkeit nicht minimiert. DiPillo (1976) betrachtet den Fall zweier multivariat normalverteilter Populationen mit gleichen A-priori-Wahrscheinlichkeiten und verwendet eine nicht erwartungstreue Minimum-Chiquadrat-Regel, die eine kleinere Fehlklassifikationswahrscheinlichkeit liefert. In Siotani & Wang (1977) werden die Fehlerraten für die Diskriminanzfunktionen  $W$  und  $Z$  verglichen. Nishi (1977) leitet engste obere Schranken für die größte der beiden Fehlerwahrscheinlichkeiten  $p_1$  bzw.  $p_2$  sowie die totale Fehlerwahrscheinlichkeit bei beliebigen A-priori-Wahrscheinlichkeiten her. Dieses geschieht für bekannte  $\mu_1$ ,  $\mu_2$ ,  $\Sigma_1$  und  $\Sigma_2$ , wobei  $V_1$  und  $V_2$  stetig oder diskret sein können. Für  $\Sigma_1 = \Sigma_2$  ergibt sich eine optimale lineare Prozedur, die das Infimum aller oberen Schranken annimmt. Falls  $\mu_1$ ,  $\mu_2$ ,  $\Sigma_1$ ,  $\Sigma_2$ ,  $q_1$  und  $q_2$  bekannt sind, optimieren Isii & Taga (1978) die Fehlerwahrscheinlichkeiten über alle möglichen Verteilungen  $V_1$  und  $V_2$  und alle möglichen Klassifikationsregeln.

## 2.4 Minimaxprinzip

Nicht immer erscheint es sinnvoll, die totale Fehlklassifikationswahrscheinlichkeit zu minimieren, weil dieses u.U. dazu führt, daß die Fehlerwahrscheinlichkeit für eine einzelne Population mit geringer A-priori-Wahrscheinlichkeit sehr hoch ist. Deshalb verwendet man oft eine *Minimaxregel*, bei der die maximale Fehlerwahrscheinlichkeit minimiert wird (zur Minimaxregel vgl. Kap. 4 des Bandes „Hypothesenprüfung“ aus der Enzyklopädie der Psychologie). Dieses ist der Fall, falls die Fehlerwahrscheinlichkeiten alle gleich und minimal sind. Für zwei Populationen  $P_1$  und  $P_2$  sind also Bereiche  $R_1$  und  $R_2$  zu finden mit

$$P(R_2 | P_1) = P(R_1 | P_2) = \text{Minimum}.$$

Die Minimax-Regel ordnet  $x$  der Population  $P_1$  zu für

$$f_1(x)/f_2(x) > c$$

und der Population  $P_2$  sonst. Die Konstante  $c$  läßt sich bestimmen, wenn man die Dichten  $f_1(x)$  und  $f_2(x)$  kennt. Die Kenntnis der A-priori-Wahrscheinlichkeiten ist bei Verwendung der Minimax-Regel nicht erforderlich. Für den Fall von multivariat normalverteilten Populationen mit gleichen Varianzen und Korrelationen leitet Johnson (1955) für zwei bzw. drei Populationen Zuordnungsregeln her. Für drei Populationen nimmt er dabei gleiche Kosten der Fehlklassifikation an. Das Verfahren wird illustriert anhand der Zuordnung von Studenten zu Unterrichtsgruppen aufgrund von Testwerten.

Von Clunis-Ross & Riffenburgh (1960) wird für zwei multivariat normalverteilte Populationen die Minimax-Regel als beste Trennfunktion unter den linearen Funktionen hergeleitet. In Anderson & Bahadur (1962) wird für zwei multivariat normalverteilte Populationen einmal die eine Fehlklassifikationswahrscheinlichkeit minimiert unter der Bedingung, daß die andere Wahrscheinlichkeit vorgegeben ist und zum anderen die Minimax-Lösung bestimmt. Da kein explizites Verfahren zur Bestimmung der auftretenden Konstanten angegeben wird, leiten Banerjee & Marcus (1965) Schranken für diese Konstanten her. Von Bhattacharya & Das Gupta (1964) wird die Klasse zulässiger Prozeduren für zwei univariate Exponentialverteilungen (im weiteren Sinne) untersucht. Eine Prozedur heißt dabei *zulässig*, falls es keine bessere Prozedur in dem Sinne gibt, daß diese Fehlklassifikationswahrscheinlichkeiten hat, die die entsprechenden Wahrscheinlichkeiten der diskutierten Prozedur nicht überschreiten und wenigstens einmal echt unterschreiten. Bei zulässigen Verfahren gibt es also keine Verfahren, die gleichmäßig besser sind. Im Spezialfall univariater Normalverteilungen mit Varianz 1 ist die übliche Prozedur eine Minimax-Regel. Neymark, Breydo & Durnovo (1970) geben für den Fall zweier multivariat normalverteilter Populationen eine lineare Minimax-Klassifikationsregel an, während Bunke (1967) Minimax-Regeln als Spezialfalle der sogenannten Q-optimalen Regeln untersucht. Während beim üblichen Bayes-Ansatz die A-priori-Verteilungen bekannt sein müssen, konstruieren Gupta & Huang (1975) sogenannte *Gamma-Minimax-Prozeduren*, bei denen nur bekannt ist, daß die A-priori-Verteilungen zu gewissen Klassen gehören.

## 2.5 Diskriminanzanalyse unter Nebenbedingungen und Kosten der Fehlklassifikation

Es gibt Fälle, wo es sinnvoll ist, unterschiedliche Kosten anzunehmen, je nachdem, welche Fehlklassifikation vorliegt. So ist es unterschiedlich zu beurteilen, ob ein Kranker irrtümlich der Population  $P_1$  der Gesunden oder ein Gesunder irrtümlich der Population  $P_2$  der Kranken zugeordnet wird. Es seien  $c_1$  die Kosten, die entstehen, wenn ein Individuum aus  $P_1$  falsch klassifiziert wird und  $c_2$  die Kosten, die entstehen, wenn ein Individuum aus  $P_2$  falsch klassifiziert wird. Es sollen Bereiche  $R_1$  und  $R_2$  gefunden werden, so daß

$$T = c_1 q_1 P(R_2 | P_1) + c_2 q_2 P(R_1 | P_2)$$

minimal wird. Dieses gilt für die Zuordnungsregel, bei der  $x$  der Population  $P_1$  zugeordnet wird, falls

$$f_1(x)/f_2(x) > c_2 q_2 / (c_1 q_1)$$

gilt und der Population  $P_2$  sonst. Man vergleiche hierzu Brown (1950) und Anderson (1951). Bei zwei multivariat normalverteilten Populationen mit gleicher Kovarianzmatrix  $\Sigma$  gibt Saxena (1967) eine Bayes-Lösung an.

Einen anderen Aspekt betrachten Zielezny & Dunn (1975). Sie gehen auch von multivariat normalverteilten  $V_1$  und  $V_2$  aus, klassifizieren jedes Individuum aber zunächst nur mit einer Teilmenge der Variablen. Nur falls dieses nicht gelingt, wird der Rest der Variablen ebenfalls gemessen und das Individuum aufgrund der Messungen aller Variablen klassifiziert. Dabei sollen die A-priori-Wahrscheinlichkeiten  $q_i$ , die Kosten der Fehlklassifikation und die Kosten der Variablenmessungen bekannt sein. Das vorgeschlagene zweistufige Verfahren wird für bekannte Parameter und für geschätzte Parameter mit anderen Verfahren verglichen.

Wenn man die unterschiedlichen Kosten der Fehlklassifikationen bei der Optimierung der Diskriminanzfunktion berücksichtigen will, muß man zumindest das Verhältnis  $c_2/c_1$  der Kosten kennen. Dieses ist oft schwer zu schätzen und willkürlich. Eine zweite Möglichkeit besteht darin, innerhalb jeder Population Fehlklassifikationswahrscheinlichkeiten vorzugeben, die noch tolerabel sind, und eine Zuordnungsregel zu bestimmen, die diese Nebenbedingungen (Ungleichungen) befriedigt. Dieses ist besonders wichtig bei diagnostischen Fragestellungen, bei denen unterschiedliche Fehlentscheidungen unterschiedlich zu bewerten sind.

Das Problem der Konstruktion von Diskriminanzfunktionen unter Nebenbedingungen wird bei Anderson (1969) für den Fall von  $k$  Populationen behandelt. Es wird dabei die Wahrscheinlichkeit einer richtigen Zuordnung maximiert unter  $(k^2-k)$  Nebenbedingungen

$$P(R_j | P_i) \leq c_{ij}, \quad i, j = 1, \dots, k, \quad i \neq k$$

an die Fehlklassifikationswahrscheinlichkeiten.

Man definiert

$$L_o(x) = 0, \quad L_i(x) = q_i f_i(x) - \sum_{j \neq i} d_{ji} f_j(x), \quad i = 1, \dots, k,$$

wobei die  $d_{ji}$  Funktionen der  $c_{ij}$  sind. Die  $L_i(x)$  heißen *modifizierte A-posteriori-Wahrscheinlichkeiten*. Man ordnet  $x$  der Population  $P_i$  zu, falls  $L_i(x)$  unter den  $(k+l)$  Funktionswerten der größte ist. Falls  $L_i(x)$  dieses Maximum annimmt, so ist  $x$  keiner der  $k$  Populationen zuzuordnen.

## 2.6 Quadratische Diskrimination

Falls man für den Fall von zwei multivariat normalverteilten Populationen mit ungleichen Kovarianzmatrizen  $\Sigma_1$  und  $\Sigma_2$  die optimale Diskriminanzfunktion bestimmt, für die die Fehlerrate minimal wird, so ergibt sich nach dem Vorgehen von Welch (1939) (vgl. 2.1)

$$Q(x) = \ln[f_1(x)/f_2(x)] = 0.5[\ln|\Sigma_1| - \ln|\Sigma_2| - \mu'_1 \Sigma_1^{-1} \mu_1 + \mu'_2 \Sigma_2^{-1} \mu_2] - 0.5[x'(\Sigma_1^{-1} - \Sigma_2^{-1})x - 2x'(\Sigma_1^{-1}\mu_1 - \Sigma_2^{-1}\mu_2)].$$

Man ordnet  $x$  der Population  $P_1$  zu für

$$Q(x) > \ln[q_2/q_1],$$

wobei  $q_1$  und  $q_2$  die A-priori-Wahrscheinlichkeiten seien. Die erste eckige Klammer enthält nur Parameter und ist deshalb eine Konstante. In der zweiten eckigen Klammer ist das erste Glied quadratisch in  $x$ , weil  $\Sigma_1 \neq \Sigma_2$  gilt, und das zweite Glied linear in  $x$ . Dieses führt zu der Bezeichnung *quadratische Diskrimination*.

Von Bartlett & Please (1963) wird der Fall  $\mu_1 = \mu_2$  und  $\Sigma_1 \neq \Sigma_2$  betrachtet, wobei bezüglich der Struktur der Kovarianzmatrizen verschiedene Spezialfälle betrachtet werden, so der Fall unabhängiger Variablen mit gleicher Varianz innerhalb jeder Population und der Fall von Variablen mit gleichen Varianzen und Korrelationen innerhalb der Populationen. Im letzteren Fall ergibt sich

$$Q(x) = c_1 Z_1 + c_2 Z_2 + c_3,$$

wobei  $c_1$ ,  $c_2$  und  $c_3$  Konstanten sind und

$$Z_1 = \sum x_j^2, Z_2 = (\sum x_j)^2.$$

Bei geeigneter Wahl von  $q_1$ ,  $q_2$ ,  $\sigma_1^2$  und  $\sigma_2^2$  hängt  $Q(x)$  entweder nur von  $Z_1$  oder nur von  $Z_2$  ab. Man interpretiert  $Z_1$  als Maß für die *Form* des durch den Meßwertvektor  $x$  beschriebenen Profils und  $Z_2$  als Maß für das *Niveau* des Profils.

Von Cooper (1963) wird gezeigt, daß die quadratische Diskrimination auch für breitere Klassen von Verteilungen als die multivariate Normalverteilung eine optimale Zuordnung liefert, während Day (1969) beweist, daß die quadratische Diskriminanzfunktion optimal ist für alle multivariaten Wahrscheinlichkeitsdichten, die zu der Klasse der quadratischen Exponentialverteilungen gehören. In Victor (1971) werden zwei FORTRAN-Programme für die quadratische Diskriminanzanalyse beschrieben, wobei die A-priori-Wahrscheinlichkeiten ebenso wie die Kosten bekannt oder unbekannt sein können. Eine Kombination von linearer und quadratischer Diskrimination schlägt Moore (1973 a) vor. Dazu werden die Populationen, bei denen man gleiche Kovarianzmatrizen vermutet, zu *Koalitionen* zusammengefaßt. Innerhalb der Koalitionen trennt man mit der linearen, zwischen den Koalitionen mit der quadratischen Diskriminanzfunktion. Bei bekannten Parametern berechnen Hildebrandt, Michaelis & Koller (1973) die beiden Fehlklassifikationswahrscheinlichkeiten  $P(R_2 | P_1)$  und  $P(R_1 | P_2)$  für die quadratische Diskriminanzfunktion. In McLachlan (1975 a) wird die Fehlerrate für die geschätzte quadratische

Diskriminanzfunktion für den Fall proportionaler Kovarianzmatrizen bei bekannten Parametern bestimmt und bei unbekannten Parametern eine asymptotische Entwicklung angegeben. Eine quadratische Diskriminanzanalyse für mehr als zwei Gruppen ohne Annahme von multivariaten Normalverteilungen schlägt Rioux (1975) vor, der vom Meßwertvektor  $x$  aus ein quadratisches Distanzmaß zu den Gruppenmittelwerten berechnet und  $x$  der Population mit der kleinsten Distanz zuordnet. Ein Vergleich der linearen und quadratischen Diskriminanzanalyse wird durch Huberty & Curry (1978) mittels einer Analyse von drei realen Datensätzen vorgenommen, wobei sieben Situationen unterschieden werden, je nachdem, ob es sich um gleiche oder verschiedene Kovarianzmatrizen und um zwei oder drei Populationen handelt. Beim *internalen* Vergleich, bei dem die Klassifikation der ursprünglichen Stichproben betrachtet wird, zeigt sich die quadratische Diskrimination überlegen. Jedoch ist die lineare Diskrimination gleich gut oder besser bei der *externalen* Klassifikation, bei der ein neuer Meßwertvektor zu klassifizieren ist. In Eye & Wiedl (1978) wird die quadratische Diskrimination zur Trennung von Personentypen ästhetischer Präferenz verwendet.

## 2.7 Zusammenhang zwischen Diskrimination und Regression

Schon früh wurde erkannt, daß die Herleitung der linearen Diskriminanzfunktion auch über eine multiple lineare Regression erfolgen kann, wobei die Prädiktorvariable eine Pseudovariable ist, deren Werte die Populationszugehörigkeit kennzeichnen. Man vergleiche hierzu Garrett (1943), Ladd (1966) und Hattemer (1974). Der entsprechende F-Test in der zugehörigen einfaktoriellen Varianzanalyse kann als Test auf Populationsunterschiede dienen. Das Distanzmaß  $D_s^2$  hängt dabei mit dem multiplen Korrelationskoeffizienten  $R^2$  durch die Beziehung

$$D_s^2 = \frac{R^2}{1-R^2} \cdot \frac{(n_1+n_2)(n_1+n_2-2)}{n_1 n_2}$$

zusammen.

Von Horst & Smith (1950) werden wegen dieser Zusammenhänge für die Bestimmung der linearen Diskriminanzfunktion Iterationsverfahren zur Berechnung multipler Regressionskoeffizienten vorgeschlagen, während Rulon (1951) eine geometrische Veranschaulichung anstrebt. Cramer (1967) vereinfacht einen Beweis von Healy (1965), der die Äquivalenz der Regressionskoeffizienten bei einem multiplen Regressionsproblem mit einer Pseudovariablen mit den Diskriminanzfunktionskoeffizienten für den Fall zweier Populationen zeigt. In Rao (1966b) wird darauf hingewiesen, daß es Fälle gibt, wo zwei Variablen einzeln betrachtet gut zwischen zwei Populationen diskriminieren, zusammengenommen aber keine Diskriminationswirkung haben. Ausgehend

von dem Zusammenhang zwischen Diskriminanz- und Regressionsanalyse zeigt Cramer (1975), daß dieses Paradoxon seine Entsprechung findet in der Beobachtung, daß die Signifikanz der t-Tests für einzelne Regressionskoeffizienten nichts über die Signifikanz des F-Tests für die gesamte Regressionsfunktion besagt. Dieser Zusammenhang wird für die Regressionsrechnung genauer untersucht.

## 2.8 Verfahren für mehrere Populationen

### 2.8.1 *Multiple Diskriminanzanalyse*

Die *multiple Diskriminanzanalyse* oder *Methode der kanonischen Variablen* ist wohl das Diskriminanzanalyseverfahren, das am häufigsten in der Psychologie verwendet worden ist, um Beobachtungsvektoren einer aus k Populationen zuzuordnen. Das Verfahren wird von Rao & Slater (1949) und Eysenck & Claridge (1962) zur Klassifikation von Neurotikern, von Power, MacRae & Muntz (1974) zur Trennung von echten und simulierenden Neurotikern, von Eysenck (1955) zur Diskrimination von Normalen, Neurotikern und Psychotikern, von Harper (1950 a, 1950 b) zur Klassifikation von Schizophrenen und von Schmidt & Cattell (1972) sowie von Overall, Higgins & Schweinitz (1976) zur Trennung psychiatrischer Diagnosegruppen verwendet. MacFadyen (1975) referiert verschiedene Arbeiten über die Klassifikation von Depressiven mit Hilfe der Diskriminanzanalyse, während Webster (1952) das Verfahren auf drei TAT-Variablen anwendet. Kelly, Veldman & McGuire (1964) versuchen Kriminelle und Schulversager zu erkennen, Petersen & Hart (1978) klassifizieren schwer erziehbare Kinder, Leton & Anderson (1964) trennen verschiedene Schulklassen, während Mallinger (1977) Lernschwächen klassifiziert. In Stahmann & Wallen (1966), Baggalay & Campbell (1967) sowie Dunteman (1966) werden Studienanfänger in bezug auf die gewählten Studienrichtungen und in Goldman & Warren (1973) sowie Zedeck (1971) bezüglich der Studienleistungen klassifiziert. Von McKeachie & Lin (1975) sowie Ware & Williams (1977) werden Unterrichtsstile von Dozenten analysiert, von Malgady (1977) werden im Rahmen der Psycholinguistik Aussagen klassifiziert, von Hand & LaFollette (1973) werden Verhaltensweisen in simulierten Spielsituationen untersucht, während Eye & Hüssy (1979) die mit Hilfe einer Clusteranalyse erhaltenen Gruppen mit unterschiedlichem Unfallrisiko trennen. Schließlich differenziert Dudzinski (1977) mit dem Verfahren Verhaltensmuster bei Kaninchen.

Beschreibungen des auf Rao (1948, 1949, 1952, 1954a, 1963, 1964) zurückgehenden Verfahrens findet man außer in den Arbeiten von Rao bei Bryan (1951), Tiedeman (1951), Horst (1956a, 1956b), Pickrel (1958), Anderson (1966) und Rulon, Tiedeman, Tatsuoka & Langmuir (1967). Es handelt sich bei dieser Methode um eine mögliche Erweiterung des Verfahrens von Fisher

(1936) von zwei Populationen auf  $k > 2$  Populationen. Es sei  $B$  die Kovarianzmatrix „zwischen den Gruppen“ und  $W$  die Kovarianzmatrix „innerhalb der Gruppen“ - analog zur Varianzanalyse. Bei bekannten Parametern hat man

$$B_T = \sum_{i=1}^k (\mu_i - \bar{\mu}) (\mu_i - \bar{\mu})' / k, W_T = \Sigma$$

mit der gemeinsamen Kovarianzmatrix  $\Sigma$  und dem Mittelwert  $\bar{\mu}$  der  $k$  Erwartungswerte. Bei Schätzung der Parameter ergibt sich

$$B_S = \sum_{i=1}^k (\bar{x}_{i..} - \bar{x}_{..}) (\bar{x}_{i..} - \bar{x}_{..})' / k, W_S = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i..}) (x_{ij} - \bar{x}_{i..})' / (n - k),$$

wobei  $n$  den Gesamtstichprobenumfang und ein Punkt die Summation über den dadurch ersetzen Index bezeichnet. Es wird ein Vektor  $a$  gesucht, so daß

$$c = a' B a / a' W a$$

maximal wird, d.h. man will die Differenzen zwischen den Gruppen maximieren und die Differenzen innerhalb der Gruppen minimieren. Dieses Maximum wird erreicht für die Eigenvektoren zu der Matrix  $W^{-1}B$ , d.h. für Lösungsvektoren  $a$  von

$$(B - c W) a = 0.$$

Bei multivariat normalverteilten Meßwertvektoren  $x$  ist  $y = a' x$  univariat normalverteilt. Verwendet man  $r$  Eigenvektoren, so erhält man  $r$  lineare Diskriminfunktionen, die man als Vektor

$$y' = (a'_1 x, \dots, a'_r x)$$

mit Koeffizientenvektoren  $a_1, \dots, a_r$  schreiben kann. Die Transformation des Erwartungswertvektors bzw. des Mittelwertvektors führt zu

$$v'_i = (a'_1 \mu_i, \dots, a'_r \mu_i) \text{ bzw. } v'_i = (a'_1 \bar{x}_i, \dots, a'_r \bar{x}_i).$$

Man bestimmt diejenige Population  $P_i$  für die

$$(y - 0.5 v_i)' v_i$$

maximal wird und ordnet  $x$  dieser Population zu.

Von Lohnes (1961) werden verschiedene Tests auf Gleichheit der Erwartungswerte bzw. Varianzen bei der multiplen Diskriminanzanalyse diskutiert und mit Hilfe von Simulationsstudien verglichen. In Lissitz & Henschke-Mason (1972) wird versucht, unter Verwendung eines Markov-Modells die Trefferra-

te in der multiplen Diskriminanzanalyse zu erhöhen. Rogers & Linden (1973) vergleichen die Ergebnisse einer multiplen Diskriminanzanalyse mit denen einer Clusteranalyse für eine Studentenstichprobe. In Geisser (1973) wird gezeigt, daß dieselbe Menge von linearen Diskriminanzfunktionen jedes Abstandsmaß zwischen den Populationen maximiert, das eine monoton wachsende Funktion der von Null verschiedenen Eigenwerte  $c$  von  $W^{-1}B$  ist. Um geeignete Variablenteilmengen zu erhalten, führt McKay (1977) schrittweise simultane Testprozeduren durch mit kontrolliertem Fehler 1. Art. Für die Bestimmung der Eigenwerte gibt Chikuse (1978) asymptotische Entwicklungen an. Weitere Versionen und Optimalitätseigenschaften der multiplen Diskriminanzanalyse geben Hudlet & Johnson (1977) an. Dabei ist das Ziel, Transformationen der Datenvektoren auf ein-, zwei- oder dreidimensionale Räume zu finden, durch die eine bessere Interpretation der Zusammenhänge zwischen den Populationen ermöglicht wird sowie ein leichteres Auffinden von Ausreißern.

### **2.8.2 Minimierung des erwarteten Verlustes**

In Lachenbruch (1975) wird ein auf Anderson (1951) zurückgehendes optimales Verfahren beschrieben, das eine Verallgemeinerung des Vorgehens von Hoel & Peterson (1949) darstellt. Es mögen  $k$  Populationen  $P_i$  mit Dichten  $f_i(x)$  und A-priori-Wahrscheinlichkeiten  $q_i$  vorliegen,  $i = 1, \dots, k$ . Falls man ein  $x$  aus  $R_i$  der Population  $P_i$  zuordnet, ergeben sich die Fehlklassifikationswahrscheinlichkeiten  $P(R_j | P_i)$  und die A-posteriori-Wahrscheinlichkeit einer Zuordnung zu  $P_i$ , falls  $x$  gegeben ist, zu

$$P(P_i | x) = q_i f_i(x) / \sum q_l f_l(x).$$

Ordnet man  $x$  der Population  $P_j$  zu, so ist der erwartete Verlust gleich

$$\sum_{i \neq j} q_i f_i(x) c_{ji} / \sum_l q_l f_l(x)$$

mit den Kosten  $c_{ji}$  bei einer Fehlklassifikation einer Beobachtung aus  $P_i$  zu  $P_j$ . Der Ausdruck wird minimal, falls der Zähler minimal wird. Dabei ist der Verlust, falls man  $x$  der Population  $P_j$  zuordnet, kleiner als der bei einer Zuordnung zu  $P_m$ , falls

$$\sum_{i \neq j} q_i f_i(x) c_{ji} < \sum_{i \neq m} q_i f_i(x) c_{mi}$$

gilt. Ist dieses für alle  $m \neq j$  erfüllt, so wird  $x$  der Population  $P_j$  zugeordnet. Speziell für den Fall, daß alle  $c_{ji}$  gleich sind, wird  $x$  zu  $P_j$  zugeordnet, falls

$$q_j f_j(x) > q_m f_m(x) \text{ für alle } m \neq j.$$

Hier von gehen Hoel & Peterson (1949) aus. Nimmt man an, daß die  $f_j$  die Dichten von multivariaten Normalverteilungen mit Erwartungswertvektoren  $\mu_j$  und gleichen Kovarianzmatrizen  $\Sigma$  sind, so wird  $x$  zu  $P_j$  zugeordnet, falls

$$\ln q_j + (x - 0.5\mu_j)' \Sigma^{-1} \mu_j$$

unter den  $k$  Gruppen maximal wird für  $j = i$ . In Guseman, Peters & Walker (1975) werden bei bekannten Parametern Optimalitätseigenschaften des Verfahrens untersucht.

### 2.8.3 Distanzmaße

Von verschiedenen Autoren werden unterschiedliche Arten von Distanzmaßen für Verteilungen definiert, um Populationen mit Hilfe der Diskriminanzanalyse trennen zu können. Insbesondere das Maß  $D^2$  von Mahalanobis wird bei Rao (1954b) diskutiert, während Matusita (1964) und Balakrishnan & Sanghvi (1968) neue Distanzmaße einführen. Falls  $k$  multivariate Normalverteilungen vorliegen mit unbekannten  $\mu_1, \dots, \mu_k$  und gemeinsamer Kovarianzmatrix  $\Sigma$ , so verwendet Srivastava (1967a) die Distanz  $D^2$ , um festzustellen, von welcher der Populationen der Meßwertvektor  $x$  den kleinsten Abstand hat. Für multivariate Normalverteilungen mit u.U. verschiedenen Kovarianzmatrizen leitet Matusita (1967) unter dem Gesichtspunkt der Distanz Diskriminfunktionen her. In Matusita (1971) wird die sogenannte *Affinität* von Verteilungen als Distanzmaß definiert. Diese ist gleich dem Integral über dem geometrischen Mittel aus den  $k$  Dichten. Dieses Maß führt zu der optimalen Regel aus 2.8.2. Für die Fehlerrate werden obere und untere Schranken durch die Affinität ausgedrückt. Glick (1973 b) vergleicht das Verhalten der Affinität und anderer Distanzmaße. Als *Klassifikationsdivergenz* bezeichnen Cleveland & Lachenbruch (1974) die Wahrscheinlichkeit, ein Individuum richtig in eine der Populationen zu klassifizieren. Die Eigenschaften dieses Distanzmaßes werden für verschiedene Verteilungsannahmen untersucht.

### 2.8.4 Andere Verfahren

In Mises (1945) wird eine Zerlegung des Raumes der Meßwertvektoren in  $k$  disjunkte Bereiche  $R_1, \dots, R_k$  angestrebt derart, daß die kleinste Wahrscheinlichkeit  $P(R_i | P_i)$  einer richtigen Zuordnung maximiert wird. In Anschluß daran maximiert Aoyama (1950) die totale Wahrscheinlichkeit

$$\sum_{i=1}^k q_i P(R_i | P_i)$$

einer richtigen Zuordnung. Voraussetzung ist, daß man die multivariaten Dichten  $f_1(x), \dots, f_k(x)$  kennt. Das Maximum ergibt sich für den Fall

$$P(R_1 | P_1) = \dots = P(R_k | P_k).$$

Diese Bedingung erlaubt auch die Bestimmung von  $R_1, \dots, R_k$ .

Von Ellison (1962) werden unter Verwendung von Stichprobenschätzungen zwei Zuordnungsregeln verglichen und als zulässig nachgewiesen, d.h. es gibt keine gleichmäßig besseren. Die eine Regel ordnet  $x$  der Population mit der minimalen quadrierten Distanz zu (minimum distance rule), während die andere Regel eine Dichte bezüglich der Parameter maximiert (restricted maximum likelihood rule). Für multivariate Normalverteilungen mit gleicher und bekannter Kovarianzmatrix zeigt Ellison (1965), daß die Minimal-Distanz-Regel äquivalent mit einer Maximum-Likelihood-Regel und einer Bayes-Regel ist. Falls man zusätzlich noch vorgegebene lineare Einschränkungen für  $\mu_1, \dots, \mu_k$  zuläßt, so zeigt Srivastava (1967b), daß die Maximum-Likelihood-Regel zulässig ist. Schließlich beschreiben Wani & Kabe (1972) in Verallgemeinerung eines Ansatzes von Uematsu (1964) ein Verfahren, das bei  $k$  Populationen  $m$  lineare Diskriminanzfunktionen so bestimmt, daß die Wahrscheinlichkeit einer Fehlklassifikation minimiert wird. Porebski (1966b) verwendet ein Verfahren, das zu  $(k-1)$  unabhängigen linearen Diskriminanzfunktionen führt, um Studenten in bezug auf ihren Studienerfolg zu klassifizieren.

### 2.8.5 Methodenvergleich

Lachenbruch (1973, 1975) vergleicht die Verfahren aus 2.8.1 und 2.8.2 unter der Annahme multivariater Normalverteilungen, gemeinsamer Kovarianzmatrizen und gleicher A-priori-Wahrscheinlichkeiten

$$q_i = 1/k, i=1, \dots, k.$$

Es werden zwei Extremfälle betrachtet, je nachdem ob die Vektoren  $\mu_1, \dots, \mu_k$  parallel sind oder ein regelmäßiges Simplex bilden. Für  $k = 3$  bedeutet dieses z.B., daß entweder  $\mu_1, \mu_2$  und  $\mu_3$  die gleiche Richtung haben oder daß sie ein gleichseitiges Dreieck bilden. In Simulationsstudien zeigte sich, daß für parallele  $\mu_1, \dots, \mu_k$  das Verfahren aus 2.8.1 ebenso gut wie das Verfahren aus 2.8.2 ist, daß das letztere Verfahren aber bei starker Verletzung der Parallelitätsvoraussetzung deutlich überlegen ist. Ferner ergab sich, daß eine Erhöhung des Stichprobenumfangs und der Anzahl der Populationen die scheinbare Wahrscheinlichkeit für eine richtige Klassifikation erniedrigt, während eine Erhöhung der Variablenzahl diese erhöht.

Das Verhalten des Verfahrens aus 2.8.2 und der entsprechenden quadratischen Diskrimination für den Fall ungleicher Kovarianzmatrizen untersucht Michael-

lis (1973a) anhand von Simulationsstudien für verschiedene Arten von Fehlerraten. Wegen der Verwendung ungleicher Kovarianzmatrizen ergibt sich, wie erwartet, eine Überlegenheit der quadratischen Diskrimination. Die Äquivalenz des Verfahrens aus 2.8.1 mit dem Verfahren, bei dem man  $x$  der Population  $P_i$  zuordnet für die

$$(\mu_i' \Sigma^{-1})x - 0.5\mu_i' \Sigma^{-1} \mu_i$$

bzw. das Stichprobenanalogen dieses Diskriminanzscores maximal wird, zeigen Kshirsagar & Arseven (1975).

## 2.9 Logistische Diskrimination

Nach dem Bayes-Ansatz ist die A-posteriori-Wahrscheinlichkeit für ein Vorliegen der Population  $P_i$  bei gegebenem Meßwertvektor  $x$  gleich

$$P(P_i | x) = q_i f_i(x) / \sum_{j=1}^k q_j f_j(x).$$

Dabei wird  $x$  der Population  $P_i$  zugeordnet, falls

$$P(P_i | x) \geq P(P_j | x) \text{ für } j=1, \dots, k$$

gilt. Falls  $f_i(x)$  die Dichte einer multivariaten Normalverteilung ist mit Erwartungswertvektor  $\mu_i$  und Kovarianzmatrix  $\Sigma$ , so lassen sich die  $P(P_i | x)$  schreiben als

$$P(P_i | x) = \exp[a_{0i} + b_i' x] P(P_k | x), i=1, \dots, k-1$$

$$P(P_k | x) = 1 / (1 + \sum_{i=1}^{k-1} \exp[a_{0i} + b_i' x]).$$

Dabei sind die Vektoren  $b_1, \dots, b_{k-1}$  und  $a'_0 = (a_{01}, \dots, a_{0(k-1)})$  bekannte Funktionen von  $\mu_1, \dots, \mu_k$  und  $\Sigma$ . Funktionen der obigen Form für  $P(P_1 | x), \dots, P(P_k | x)$  mit unbekannten Parametervektoren heißen *multivariate logistische Funktionen*. Derartige Funktionen treten nicht nur auf für multivariate Normalverteilungen mit gleichen Kovarianzmatrizen, sondern auch für unabhängige dichotome Variablen, für dichotome Variablen in einem loglinearen Modell mit gleichen Effekten zweiter und höherer Ordnung und in Mischungen dieser letzten Verteilungsform mit multivariaten Normalverteilungen mit gleichen Kovarianzmatrizen.

Eine der ersten ausführlichen Beschreibungen der Verwendung der multivariaten logistischen Funktion in der Diskriminanzanalyse gibt Cox (1966). Von

Day & Kerridge (1967) werden Maximum-Likelihood-Schätzungen für die Parametervektoren angegeben und die Vielzahl der Modelle diskutiert, die unter die logistische Diskrimination fallen. Sie zeigen, daß, falls die Ausgangsstichprobe im Prinzip ohne Fehler trennbar ist, dieses von der durch die logistische Diskrimination erhaltenen Zuordnungsregel geleistet wird. In Anderson (1972) erfolgt die Anwendung der logistischen Diskrimination auf qualitative Daten und die Erweiterung auf mehr als zwei Populationen. Bei Vorliegen von  $k$  Populationen untersucht Anderson (1973) die Schätzung der logistischen Parameter für den Fall einer Mischstichprobe, den Fall getrennter Stichproben und den Fall unbekannter Mischungsanteile. Wie Anderson (1975) diskutiert, ergeben sich lineare Diskriminanzfunktionen, falls die Abhängigkeitsstrukturen für alle  $k$  Populationen gleich sind. Andernfalls führt der logistische Ansatz zu quadratischen Diskriminanzfunktionen mit einer großen Zahl von Parametern. Dabei ergeben sich numerische Schwierigkeiten bei der iterativen Schätzung dieser Parameter. Da außerdem die Gefahr einer Überanpassung besteht, betrachtet Anderson (1975) Approximationen mit einer verringerten Parameterzahl. Da bei kleinen Stichproben die Maximum-Likelihood-Schätzungen der logistischen Parameter eine erhebliche Verzerrung (bias) zeigen, führen Anderson & Richardson (1979) eine Korrektur ein, deren Effizienz durch eine Simulationsstudie überprüft wird. Von Anderson (1979) wird eine allgemeine Maximum-Likelihood-Schätzprozedur für die Anteile verschiedener Populationen an einer Mischpopulation angegeben. Insbesondere ergibt sich damit eine Möglichkeit, durch Einbeziehung nicht sicher zugeordneter Individuen die Schätzungen von Diskriminanzfunktionen zu verbessern.

## 2.10 Kovariate Diskriminanzanalyse

Ausgangspunkt sei das Problem der Diskrimination von zwei Populationen, wenn bekannt ist, daß diese sich bezüglich einer gegebenen Teilmenge von Variablen nicht unterscheiden. Solche Variablen heißen *Nebenvariablen* (concomitant variables) im Unterschied zu den *Hauptvariablen*, in denen sich die Populationen unterscheiden. Wenn die Nebenvariablen mit Hauptvariablen korreliert sind, können sie zur Diskrimination beitragen, obwohl sie selbst nicht diskriminieren. Dieser theoretische Vorteil einer Berücksichtigung der Nebenvariablen durch das Verfahren einer Kovarianzanzpassung kann in der Praxis verlorengehen, falls die Korrelationen unbekannt sind und aus den Daten geschätzt werden müssen.

In der Bezeichnungsweise von Rao (1966b) sei  $Y$  ein  $p$ -dimensionaler Zufallsvektor der Hauptvariablen und  $Z$  ein  $q$ -dimensionaler Zufallsvektor der Nebenvariablen. Mit  $E_1$  bzw.  $E_2$  werde die Erwartungswertbildung bezüglich zweier  $(p+q)$ -dimensionaler Normalverteilungen mit gleichen Kovarianzma-

trizen und möglicherweise verschiedenen Erwartungswertvektoren verstanden. Von Interesse sind die vier Hypothesen

$$H_{01} : E_1[Y] = E_2[Y] \text{ und } E_1[Z] = E_2[Z]$$

$$H_{02} : E_1[Y|Z] = E_2[Y|Z]$$

$$H_{03} : E_1[Y] = E_2[Y] \text{ unter Voraussetzung von } E_1[Z] = E_2[Z]$$

$$H_{04} : E_1[Y] = E_2[Y].$$

Es sei  $D_{p+q}^2$  bzw.  $D_p^2$  bzw.  $D_q^2$  die geschätzte Mahalanobis-Distanz basierend auf  $(Y, Z)$  bzw. auf  $Y$  bzw. auf  $Z$ . Ferner sei  $c = n_1 n_2 / N$  mit  $N = n_1 + n_2$ . Dann schlägt Rao (1966b) für die obigen Hypothesen die Teststatistiken

$$T_1 = \frac{c(N-p-q-1)}{(p+q)(N-2)} D_{p+q}^2, T_2 = \frac{N-p-q-1}{p} - \frac{c(D_{p+q}^2 - D_q^2)}{N-2+cD_q^2},$$

$$T_3 = \frac{c}{N-2} (D_{p+q}^2 - D_q^2), T_4 = \frac{c(N-p-1)}{p(N-2)} D_p^2$$

vor. Von Cochran & Bliss (1948) und Cochran (1964b) wird  $T_2$  sowohl für  $H_{02}$  als auch für  $H_{03}$  verwendet, obwohl Rao (1949) zeigt, daß es etwa besser ist,  $T_3$  für  $H_{03}$  zu verwenden. In Narain (1950) wird bemerkt, daß Tests zum Trennen von Populationen bei Vorliegen von Nebenvariablen trennschärfster sind. Jedoch müssen bei der Stichprobenerhebung dann auch die Nebenvariablen entsprechend berücksichtigt werden. Von Cochran (1964b) wird das Verfahren der kovarianzangepaßten Variablen mit dem Verfahren verglichen, bei dem die Nebenvariablen als Diskriminatoren in die Diskriminanzfunktion miteingenommen werden. Das erste Verfahren liefert zwar trennschärfere Tests, jedoch gibt es kaum Unterschiede in der Klassifikationsgüte. In Rao (1966b) werden außer den Tests die Schätzung der Diskriminanzfunktion und der Effekt einer Erhöhung der Variablenzahl untersucht.

Die Kovarianzanpassung erfolgt so, daß man die Regressionsmatrix  $B$  aus der Stichprobe schätzt und anstelle des Meßwertvektors  $x$  den mit Hilfe der Schätzung von  $B$  aus  $x$  erhaltenen angepaßten Vektor  $x^*$  verwendet. Einsetzen von  $x^*$  anstelle von  $x$  in die Diskriminanzfunktion  $W$  (vgl. 2.1) ergibt  $W^*$ . Von Memon & Okamoto (1970) wird eine asymptotische Entwicklung für die Verteilung von  $W^*$  und beide Fehlklassifikationswahrscheinlichkeiten angegeben. Von Lachenbruch (1977) werden für eine Haupt- und eine Nebenvariable die Formen der Klassifikationsbereiche untersucht.

## 2.11 Sequentielle Diskrimination

Falls der Abstand zwischen den Populationen klein ist, ist es u.U. schwer, ein Individuum aufgrund eines Vektors  $x$  einer der Populationen zuzuordnen.

Falls es dann möglich ist, von einem Individuum mehrere *unabhängige* Meßwertvektoren zu erheben, so kann man an ein *sequentialles* Verfahren denken. Dazu erhebt man in zeitlicher Aufeinanderfolge neue Vektoren, bis man aufgrund einer Stoppregel das Individuum einer der Populationen zuordnen kann. Bei zwei Populationen  $P_1$  bzw.  $P_2$  seien  $e_1$  bzw.  $e_2$  die maximalen Fehleranteile, die man in  $P_1$  bzw.  $P_2$  zulassen will und

$$A = (1 - e_2)/e_1, \quad B = e_2/(1 - e_1).$$

Von Wald (1947) werden *Likelihood-Quotienten-Sequentialtests* angegeben, die sich auch in der Diskriminanzanalyse verwenden lassen. Man bildet dazu für die bekannten Dichten  $f_i(W_T(x))$  der Diskriminanzfunktion  $W_T(x)$  (vgl. 2.1) das Produkt der Dichtequotienten

$$Q_i = \prod_{j=1}^i \frac{f_2(W_T(x_j))}{f_1(W_T(x_j))}, \quad i = 1, 2, \dots,$$

falls man im  $i$ -ten Schritt den Vektor  $x_i$  erhalten hat, beginnend beim ersten Vektor  $x_1$ . Falls im  $i$ -ten Schritt  $Q_i \leq B$  gilt, so ordnet man das Individuum der Population  $P_1$  zu, falls  $Q_i \geq A$  gilt, so ordnet man das Individuum  $P_2$  zu. In diesen Fällen bricht das Verfahren ab. Für  $B < Q_i < A$  erhebt man den  $(i+1)$ -ten Meßwertvektor und führt die entsprechende Überprüfung durch, usw. Unbekannte Verteilungsparameter kann man mit der Maximum-Likelihood-Methode schätzen. Die Kenntnis der A-priori-Wahrscheinlichkeiten ist nicht erforderlich, da die Fehlklassifikationswahrscheinlichkeiten vorgegeben werden.

Von Armitage (1950) wird die Sequentialversion der multiplen Diskriminanzanalyse für  $k$  Populationen betrachtet. Mallows (1953) geht davon aus, daß prinzipiell unendlich viele Komponenten des Meßwertvektors  $x$  vorliegen und mißt sequentiell immer mehr Variablen, bis er eine Entscheidung für  $P_1$  oder  $P_2$  erhält. Dabei verwendet er anstelle von  $Q_i$  den Quotienten der multivariaten Dichten

$$Q_{im} = f_2(x_1, \dots, x_i) / f_1(x_1, \dots, x_i), \quad i=1, 2, \dots$$

Auch das Problem der optimalen Reihenfolge der zu messenden Variablen wird diskutiert. Ein nichtparametrisches Verfahren für die gleiche Fragestellung, das auf Ordnungsstatistiken beruht, gibt Kendall (1966) an. Da hierbei u.U. ein erheblicher Anteil von Individuen nicht klassifiziert wird, schlägt Richards (1972) vor, bei jedem Schritt die Verteilung aller Variablen mitzubücksichtigen. Falls zwei multivariate Normalverteilungen mit gleicher Kovarianzmatrix vorliegen und die A-priori-Wahrscheinlichkeiten gleich sind, zeigen Smith & Yau (1972), daß eine lineare sequentielle Diskriminanzfunktion, die nach der Methode der kleinsten Quadrate erhalten wird, mit der Likelihood-Quotienten-Diskriminanzfunktion äquivalent ist.

Von Samuel (1963) werden für die ursprüngliche Fragestellung und den Fall von zwei vollständig bekannten Populationen Klassen von optimalen sequentiellen Entscheidungsregeln spezifiziert. Im Falle multivariater Normalverteilungen bestimmt Geisser (1966) die sequentiellen A-posteriori-Wahrscheinlichkeiten. Von Freedman (1967) wird der Fall abzählbar unendlich vieler Populationen betrachtet. Daß der Likelihood-Quotienten-Test in diesem Fall nicht immer zu sinnvollen Ergebnissen führt, zeigt Yarbrough (1971).

Man kann auch den Fall betrachten, daß für jede der beiden Populationen eine eindeutig zugeordnete Stichprobe von je  $n$  Individuen vorliegt und man eine neue Stichprobe von  $n$  Individuen einer der beiden Populationen zuordnen möchte, wobei man diese neue Stichprobe sequentiell erhebt. Unter Verwendung sequentieller Ränge und des Likelihood-Quotienten-Tests schlagen Fu & Chien (1967) ein nichtparametrisches Verfahren vor. Für den Fall eindimensionaler stetiger Verteilungen wird von Woinsky & Kurz (1969) ein anderes nichtparametrisches Verfahren, das auf der Differenz zweier Mann-Whitney-Statistiken beruht, diskutiert. Hierbei wird die Stoppregel dadurch festgelegt, daß man das Maximum der beiden Fehlklassifikationswahrscheinlichkeiten vorgibt.

## 2.12 Zeitreihen

Auch für das Problem der Diskriminanzanalyse von unendlich vielen abhängigen Daten sind für den Fall spezieller Abhängigkeitsstrukturen Diskriminanzverfahren entwickelt worden. So gibt Rao (1951) sequentielle Verfahren für die Trennung von Gleitmittelprozessen und autoregressiven Prozessen an. Von Kashyap (1978) wird eine optimale Entscheidungsregel konstruiert, die die Fehlerrate minimiert, um eine Zeitreihe einem von mehreren autoregressiven Prozessen zuzuordnen.

Rao & Varadarajan (1963) nehmen an, daß es möglich ist, an einem Individuum eine große Anzahl von Hilfsbeobachtungen zu machen, die einem Gauß-Prozeß unterliegen, d.h. für die die endlich-dimensionalen Randverteilungen multivariat normal sind. Es wird die Frage untersucht, ob die Diskrimination zweier Gauß-Prozesse perfekt wird, wenn man immer mehr Beobachtungen einbezieht, oder ob sich die Fehlklassifikationswahrscheinlichkeiten stabilisieren. In Tallis (1970) wird ein optimales Diskriminationsverfahren entwickelt, das für diesen Fall die erwarteten Kosten der Fehlklassifikation minimiert. Den Fall zweier stationärer Gauß-Prozesse betrachtet Bandyopadhyay (1979). Speziell für gleiche Fehlklassifikationswahrscheinlichkeiten untersucht er den Einfluß der Populationsverteilungen auf die Wahrscheinlichkeit für eine richtige Klassifikation. In Zusammenhang mit den Untersuchungen über Gauß-Prozesse steht auch die Arbeit von Cacoullos (1973b) über den normalen p-

dimensionalen Diffusionsprozeß (Wiener-Prozeß mit Drift), der auf den Fall der endlich-dimensionalen Normalverteilung zurückgeführt werden kann. Es wird diskriminiert mit Hilfe von Minimal-Distanz-Regeln und es werden untere Schranken für die Wahrscheinlichkeit einer richtigen Zuordnung gegeben.

## 2.13 Variablenauswahl

In der Praxis ist es bei einem Diskriminationsproblem meist nicht von vornherein klar, welche Variablen gute Diskriminatoren sind, d.h. viel zur Trennung der Populationen beitragen, und welche nicht. Deshalb mißt man meist alle Variablen, die problemrelevant erscheinen, und versucht nachträglich diejenigen zu eliminieren, die wenig zur Trennung beitragen. Bei zwei multivariat normalverteilten Populationen sind die Einzelvariablen als univariat normalverteilt anzusehen, und Lachenbruch (1975) schlägt in diesem Falle vor, für jede Einzelvariable einen zweiseitigen t-Test für unabhängige Stichproben zu rechnen und nur Variablen mit signifikantem Testergebnis zu berücksichtigen. Da es sich um p abhängige Tests handelt, wird dabei eine Alphaadjustierung nach dem Bonferroni-Verfahren vorgenommen, d.h. es wird jeweils bei  $\alpha/p$  beurteilt. Offensichtlich darf die Anzahl p der Variablen nicht zu groß sein. Ein anderer Nachteil des Verfahrens ist, daß die Trennschärfe des t-Tests mit den Stichprobenumfängen  $n_1$  und  $n_2$  wächst und bei hohen Stichprobenumfängen auch geringe Populationsunterschiede in einer Variablen zu einem großen Wert  $t_i$  der Statistik des t-Tests für die Variable i führen können. Dieses ergibt sich auch aus der Berechnung der Mahalanobis-Distanz

$$D_{Si}^2 = t_i^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)$$

der Populationen für die Variable i. Diese ist u.U. klein, falls  $n_1$  und  $n_2$  groß sind, d.h. die Variable i hat dann nur geringe Trennschärfe.

Wegen der Abhängigkeit der Variablen kann es sein, daß das trennschärfste Variablenpaar die trennschärfste Einzelvariable nicht enthält. Allgemein enthält die trennschärfste Teilmenge von r Variablen nicht unbedingt die trennschärfste Teilmenge von (r-1) Variablen. Es wäre naheliegend, für alle  $(2^{P-1})$  nichtleeren Variablenteilmengen jeweils die Distanz  $D_S^2$  zu berechnen und für diese entsprechende Signifikanztests durchzuführen. Ein Computerprogramm für die Analyse aller Variablenteilmengen beschreibt McCabe (1975). Dieses Programm wird mit dem Programm BMD 07M (vgl. Dixon, 1976) für eine schrittweise Diskrimination verglichen. Linhart (1961) schlägt vor, eine Variablenteilmenge von r Variablen nur dann zur Diskrimination zu verwenden, wenn die Fehlerrate für alle Teilmengen mit weniger als r Variablen größer ist. Urbakh (1971) gibt Schätzungen für den Verlust an Trennschärfe an, wenn man jeweils eine Variable wegläßt. Als Alternative zu  $D_S^2$  schlagen Wolde-

Tsadik & Yu (1979) die *Konkordanzwahrscheinlichkeit* vor, d.h. die Wahrscheinlichkeit, daß eine Variablenmenge und eine Teilmenge von ihr ein Individuum in dieselbe Population klassifizieren. Diese Wahrscheinlichkeiten werden dann nach inhaltlichen Gesichtspunkten beurteilt. Von McKay (1978) wird ein graphisches Verfahren zur Variablenauswahl, das in der multiplen Regressionsanalyse verwendet wird, adaptiert.

Am gebäuchlichsten sind wohl die Verfahren der schrittweisen Diskrimination. Man unterscheidet dabei Verfahren der *Vorwärtsauswahl*, bei der man schrittweise bei einer Variablen beginnend immer eine Variable hinzunimmt, und der *Rückwärtsauswahl*, bei der man vom vollen Variablenatz ausgehend schrittweise Variablen wegläßt. Bei Habbema, Hermans & Broek (1974) wird ein Verfahren zur Vorwärtsauswahl für stetige Variablen beschrieben, das auf einer Schätzung der unbekannten Verteilungsdichten beruht. Zunächst wird diejenige Einzelvariable gesucht, die als Diskriminator minimale Kosten verursacht. Im zweiten Schritt wird diejenige Variable hinzugenommen, die mit der ersten Variablen zusammen unter allen Paaren, die die erste Variable enthalten, die Kosten minimiert. Zu diesem Paar wird entsprechend eine dritte Variable gesucht, usw. Henschke & Chen (1974) berechnen zunächst die Diskriminanzfunktion für alle p Variablen und wählen als erste Variable diejenige mit der höchsten Korrelation bezüglich der Diskriminanzfunktion. Dann bestimmen sie die partiellen Korrelationen der übrigen Variablen mit der Diskriminanzfunktion nach Ausschalten der ersten Variablen. Diejenige Variable mit der höchsten partiellen Korrelation wird als zweite Variable gewählt, usw. Dieses Verfahren beschreiben auch Baitsch & Bauer (1956). Verschiedene Probleme, die bei der schrittweisen Diskrimination auftreten, z.B. die Frage des benötigten Stichprobenumfangs, diskutieren Nakache & Dusserre (1975). Das Verhalten der tatsächlichen und der geschätzten Fehlerraten bei einer Vorwärtsauswahl untersucht Moran (1975a) in einer Simulationsstudie. Es zeigt sich, daß man, falls solange Variablen hinzugenommen werden wie der Zuwachs von  $D_S^2$  noch bei 30% signifikant ist, das gleiche Verhalten der so gewonnenen Teilmenge und der Gesamtvariablenmenge erwarten kann. Auch das Programm BMD 07M (Dixon, 1976) verwendet eine Vorwärtsprozedur. Eine Variable wird hinzugenommen, falls sie eine gewisse Statistik maximiert und der Wert dieser Statistik über einem gegebenen Schwellenwert liegt. Von der vierten Variablen an wird bei einem Wert der Statistik unterhalb der Schwelle die Variable mit dem niedrigsten Kriteriumswert aus der Variablenmenge eliminiert, wobei die neue Variable i. a. zur Menge hinzugenommen wird. Es stehen in dem Programm drei Kriterien zur Auswahl. Ein Kriterium ist die F-Statistik zwischen den Gruppen.

Von McLachlan (1976) wird der Fall zweier multivariat normalverteilter Populationen mit gleicher Kovarianzmatrix und gleichen A-priori-Wahrscheinlichkeiten bei Verwendung der Diskriminanzfunktion W (vgl. 2.1) untersucht. Es

wird für eine weggelassene Teilmenge von Variablen ein approximativer Konfidenzkoefizient bestimmt, der das Nichtanwachsen der Fehlerrate mißt und als Indikator für den zusätzlichen Trenneffekt der weggelassenen Variablenmenge dient. Dieser Ansatz kann für Vorwärts- und Rückwärtssauswahl verwendet werden. Für qualitative Daten geben Goldstein & Dillon (1977) eine Erweiterung des Verfahrens von Hills (1967) an, die darin besteht, daß die Stopptregel auf der Informationsstatistik von Kuliback (1959) beruht.

Für die multiple Diskriminanzanalyse schlägt McKay (1977) ein Auswahlverfahren vor, in dem schrittweise simultane Testprozeduren mit kontrolliertem Alpharisiko durchgeführt werden. Mit Hilfe einer Simulationsstudie demonstrieren Hecker & Wegener (1978), daß die Hinzunahme immer weiterer Variablen in eine schrittweise Diskriminanzanalyse die Wahrscheinlichkeit dafür erhöht, daß man Variablen auswählt, die rein zufällig die vorgegebenen Populationen gut trennen. Diese Gefahr ist vor allem durch die leichte Verfügbarkeit der SPSS-Programme (Nie, Hill, Jenkins, Steinbrenner & Bent, 1975) und BMD-Programme (Dixon, 1975) gewachsen.

Einen Vergleich verschiedener älterer Auswahlverfahren geben Baitsch & Bauer (1959). Für zwei multivariat normalverteilte Populationen und unter Verwendung der linearen Diskriminanzfunktion vergleichen Weiner & Dunn (1966) mehrere Auswahlverfahren anhand von fünf medizinischen Datensätzen. Das erste Verfahren verwendet die Variablen mit den größten Werten der t-Statistik, das zweite Verfahren verwendet diejenigen mit den größten standardisierten Koeffizienten in der Diskriminanzfunktion, das dritte Verfahren fügt schrittweise jeweils die Variable hinzu, die die Restvarianz möglichst stark reduziert, während das vierte Verfahren die Variablen per Zufall auswählt. Die Güte der Verfahren wird gemessen durch den Anteil der Fehlklassifikationen bei Folgestichproben. Dabei zeigt sich, daß nur das erste und dritte Verfahren besser als das vierte Verfahren sind. In Eisenbeis, Gilbert & Avery (1973) werden für zwei empirische Datensätze fünf Verfahren, um die relative Trennschärfe der Einzelvariablen zu beurteilen, sechs Verfahren zur Auswahl von Variableneilmengen und ein Verfahren zur Beurteilung von Teilmengen verschiedenen Umfangs miteinander verglichen. Habbema & Hermans (1977) vergleichen fünf Programme zur Variablenauswahl, während Farver & Dunn (1979) eine Vorwärtsauswahl und eine modifizierte Vorwärtsauswahl, bei der zu Beginn die Variablen mit den kleinsten Werten der t-Statistik eliminiert werden, miteinander vergleichen. In einer Simulationsstudie zeigte sich die Überlegenheit der modifizierten Vorwärtsauswahl in Hinblick auf das Kriterium der Wahrscheinlichkeit einer richtigen Klassifikation. Die Vor- und Nachteile verschiedener Auswahlverfahren diskutieren auch Schaafsma & Vark (1979).

Von Melrose, Stroebel & Glueck (1970), Altman, Evenson & Cho (1976) sowie von Clavelle & Butcher (1977) wird eine schrittweise Version der mul-

tiplen Diskriminanzanalyse verwendet, um psychiatrische Populationen zu trennen. Coolidge (1976) trennt mit der schrittweisen Diskriminanzanalyse Hirnverletzte und psychiatrische Patienten, Swiercinsky & Warnock (1977) trennen Hirnverletzte und Nichthirnverletzte, während Sheslow & Erickson (1975) depressive von nichtdepressiven Studenten trennen. Bledsoe (1973) klassifiziert sechs Gruppen von Lehrern, einmal mit der schrittweisen Diskrimination und einmal mit der schrittweisen Regression und vergleicht die Ergebnisse. Walter & Porges (1976) suchen diejenigen physiologischen Variablen, die besonders empfindlich auf psychologische Manipulationen reagieren. Jedoch warnen Lachin & Schachter (1974) vor der Verwendung der schrittweisen Diskrimination, da diese fast immer signifikante Ergebnisse liefert. Zur Illustration verwenden sie EKG- und EEG-Daten unter Reiz- und Kontrollbedingung. Das Programm BMD 07M liefert auch unter Nichtreizbedingung signifikante Ergebnisse, und unter Reizbedingung ergibt sich eine übertrieben gute Trennung. Bei Verwendung eines nichtschrittweisen Verfahrens werden solche Artefakte nicht beobachtet.

### *3. Inferenzstatistik*

#### **3.1 Signifikanztests**

Nach Durchführung einer Diskriminanzanalyse kann man nach dem Erfolg der Analyse fragen. Eine Möglichkeit, diesen Erfolg zu bewerten, besteht in der Durchführung eines Signifikanztests auf Populationsunterschiede. Für zwei multivariat normalverteilte Populationen wird in 2.1 eine F-Statistik angegeben, die auf  $D_S^2$  beruht. Diese ist äquivalent mit der Statistik

$$T^2 = n_1 n_2 D_S^2 / (n_1 + n_2)$$

von Hotelling (Lachenbruch, 1975). Tests auf Populationsunterschiede bei k Populationen diskutieren u.a. Rao (1963-1964) und Lohnes (1961).

Signifikanztests werden auch verwendet, um geeignete Teilmengen von Variablen auszuwählen. So gibt Rao (1970) einen F-Test an, der auf Mahalanobis-Distanzen beruht, um zu testen, ob eine Variableneiteilmenge zwei multivariat normalverteilte Populationen trennt. Für die gleiche Problemstellung beschreibt McKay (1976) simultane Tests.

Auch Anpassungstests, um zu testen, ob eine Menge von vorgegebenen Diskriminanzfunktionen mit den Daten verträglich ist, werden angegeben, z.B. von Kshirsagar (1970, 1971) für s hypothetische Diskriminanzfunktionen für k Populationen. Von Chou & Muirhead (1979) wird die Verteilung einer Teststatistik für die Anzahl der benötigten Diskriminanzfunktionen in der multi-

plen Diskriminanzanalyse untersucht. Für zwei beliebige konkurrierende Diskriminationsverfahren beschreibt Goldstein (1976) zwei Signifikanztests, die die Nullhypothese, daß beide Verfahren nicht besser als eine Zufallszuordnung der Individuen zu den  $k$  Populationen sind, gegen die Alternativhypothese testen, daß das erste Verfahren trennschärfer als das zweite Verfahren ist.

Auch Tests dafür, ob überhaupt die Voraussetzungen für eine Diskriminanzanalyse erfüllt sind, werden vorgeschlagen. So gibt z.B. Rao (1969) einen Test dafür an, daß ein Individuum zu einer von  $k$  vorgegebenen Populationen gehört, während McDonald, Lowe, Smidt & Meister (1976) testen, ob der Meßwertvektor eines Individuums überhaupt zu einer von zwei vorgegebenen multivariaten Normalverteilungen gehören kann.

## 3.2 Schätzungen

Ein Grundproblem der Diskriminanzanalyse ist die Schätzung der Diskriminanzfunktion bzw. der Koeffizienten dieser Funktion. Für multivariat normalverteilte Populationen berechnet Das Gupta (1968) die Erwartungswerte und Kovarianzen für die Koeffizientenschätzungen und bestimmt die asymptotischen Verteilungen auch für den allgemeinen Fall nicht normalverteilter Populationen. Aufbauend auf diesen Ergebnissen konstruiert er optimale Tests für verschiedene Nullhypthesen, die die Koeffizienten der Diskriminanzfunktion betreffen. Falls nur bekannt ist, daß die vorliegende Stichprobe einer Mischung von zwei univariaten Normalverteilungen mit gemeinsamer Varianz entstammt, schätzen Ganesalingam & McLachlan (1978) die optimale Diskriminanzfunktion und bestimmen ihre asymptotische relative Effizienz bezüglich des optimalen Verfahrens bei bekannten Parametern.

Zur Bewertung der Trennschärfe eines Diskriminationsverfahrens benötigt man eine zuverlässige Schätzung der Fehlerrate, um abschätzen zu können, mit welcher Sicherheit später erhobene Individuen richtig klassifiziert werden. Eine Übersicht über die zahlreichen Lösungsansätze für dieses Problem gibt die Bibliographie von Toussaint (1974) mit 188 Literaturstellen. Neuere zusammenfassende Darstellungen geben Schaafsma & Vark (1977, 1979). Für den Fall von zwei multivariat normalverteilten Populationen untersucht McLachlan (1977b, 1978) das mit der Schätzung der Fehlerrate in Zusammenhang stehende Problem, wie sich eine Diskriminanzfunktion verhält, wenn man die Fehlklassifikationswahrscheinlichkeiten durch obere Konfidenzgrenzen einschränkt.

## 4. Robustheit

### 4.1 Lineare Diskriminanzfunktion

Unter der *Robustheit* einer Diskriminanzfunktion versteht man die Eigenschaft, auch dann noch verschiedene Populationen hinreichend gut trennen zu können, falls wesentliche Voraussetzungen bei der Herleitung der Funktion nicht erfüllt sind. Solche Voraussetzungen sind bei der linearen Diskriminanzfunktion z.B. die multivariate Normalverteilung für die Ausgangspopulationen, die Gleichheit der Kovarianzmatrizen, die richtige Klassifikation der Ausgangsstichproben und das Vorliegen vollständiger Meßwertvektoren.

Die Robustheit der linearen Diskriminanzfunktion bei Vorliegen von anderen Verteilungstypen als der Normalverteilung untersuchen Subrahmaniam & Chinganda (1978) sowie Chinganda & Subrahmaniam (1979), während Ahmed & Lachenbruch (1975) das Verhalten dieser Funktion speziell bei Vorliegen von Ausreißerverteilungen betrachten. Es zeigt sich i. a. eine große Empfindlichkeit der linearen Diskriminanzfunktion gegen Verletzungen der Normalverteilungsannahme. Aus den Simulationsstudien von Marks & Dunn (1974) sowie Wahl & Kronmal (1977) ist erkennbar, daß die lineare Diskriminanzfunktion von Fisher bei großen Stichproben und annähernd gleichen Kovarianzmatrizen ähnlich gut diskriminiert wie die quadratische Diskriminanzfunktion, dieser aber bei stark verschiedenen Kovarianzmatrizen deutlich unterlegen ist. Bei kleinen Stichproben ergibt sich kaum ein Unterschied.

Das Verhalten der linearen Diskriminanzfunktion für den Fall, daß schon die Ausgangsstichproben nicht richtig klassifiziert sind, untersucht Lachenbruch (1966, 1974, 1975). Aus Simulationsstudien ergibt sich, daß für gewisse Modellannahmen die tatsächliche Fehlerrate nur wenig beeinflußt wird, während der scheinbare Fehler und die Mahalanobis-Distanz stark betroffen sein können.

Für unvollständige Meßwertvektoren schätzen Smith & Zeis (1973) und Srivastava & Zaatar (1972) die unbekannten Parameter mit der Maximum-Likelihood-Methode. Chan & Dunn (1972) vergleichen in einer Simulationsstudie mehrere Verfahren, um mit unvollständigen Meßwertvektoren zurechtzukommen. Dazu gehören die Verwendung nur vollständiger Vektoren, die Verwendung aller Meßwerte ohne Ersetzung fehlender Werte durch Pseudomeßwerte, die Ersetzung fehlender Werte durch Mittelwerte und zwei weitere ursprünglich in der Regressionsrechnung verwendete Verfahren. Insbesondere das Verfahren der Meßwertersetzung durch Mittelwerte gehört zu den beiden besten Methoden.

Eine Übersicht über verschiedene Untersuchungen zum Verhalten der linearen Diskriminanzfunktion bei Verletzung der Voraussetzungen gibt Krzanowski

(1977). Insbesondere diskutiert er die Ergebnisse einer Simulationsstudie für den Fall, daß gleichzeitig diskrete und stetige Variablen vorliegen.

## 4.2 Quadratische Diskriminanzfunktion

Mit Simulationsstudien untersuchen Clarke, Lachenbruch & Broffitt (1979) das Verhalten der quadratischen Diskriminanzfunktion bei Vorliegen nicht-normalverteilter Zufallsvariablen. Außer im Fall sehr schiefer Verteilungen erweist sich die quadratische Diskriminanzfunktion als ziemlich robust. Während sich die totale Fehlerrate bei unterschiedlichen Verteilungsannahmen als relativ stabil zeigt, ergeben sich große Unterschiede für die Einzelfehlklassifikationswahrscheinlichkeiten.

Eine Simulationsstudie von Lachenbruch (1979) zeigt, daß der Einfluß einer Fehlklassifikation der Ausgangsstichproben einen erheblichen Einfluß auf die Fehlerrate der quadratischen Diskriminanzfunktion hat. Anders als bei der linearen Diskriminanzfunktion ergeben sich auch im Fall gleich großer Raten für die ursprüngliche Falschzuordnung keine besseren Ergebnisse.

## 4.3 Robuste Diskriminanzfunktionen

Eine mögliche Konsequenz, die man aus der Empfindlichkeit der Diskriminanzfunktionen bezüglich Verletzungen der Voraussetzungen ziehen kann, ist die Konstruktion von *robusten Diskriminanzfunktionen*. In Ahmed & Lachenbruch (1977) werden Mittelwerte und empirische Kovarianzen in der linearen Diskriminanzfunktion durch verschiedene robustere Schätzungen ersetzt, z.B. durch getrimmte Mittel. Ein zweiter Ansatz verwendet robuste Regressionsschätzungen, während ein dritter Ansatz vor Berechnung der Diskriminanzfunktion störende Meßwerte eliminiert. In einer Simulationsstudie werden die verschiedenen Ansätze für eine Reihe von Verteilungsannahmen verglichen. Eine ähnliche Studie legen Randles, Broffitt, Ramberg & Hogg (1978b) vor. Neben robusten Schätzern, die von den Originaldaten ausgehen, werden hier auch Rangverfahren betrachtet und nicht nur für die lineare, sondern auch für die quadratische Diskriminanzfunktion robustere Versionen vorgeschlagen.

Für verschiedene Verteilungsannahmen vergleichen Koffler & Penfield (1979) das Verhalten der linearen Diskriminanzfunktion, der quadratischen Diskriminanzfunktion, eines nichtparametrischen Nächster-Nachbar-Verfahrens und dreier nichtparametrischer Dichteschätzer in einer Simulationsstudie. Als robusteste Verfahren erweisen sich das Nächster-Nachbar-Verfahren und der Loftsgaarden-Quesenberry-Dichte-Schätzer.

## 5. Nichtparametrische Verfahren

### 5.1 Nichtparametrische Zuordnungsregeln

Anstelle von robusten Verfahren kann man bei Verletzung der Normalverteilungsannahme auch nichtparametrische Verfahren verwenden. Die meisten nichtparametrischen Diskriminationsmethoden verwenden nichtparametrische Schätzungen der unbekannten Dichten und benutzen den Quotienten dieser Dichten als Diskriminanzfunktion, analog zu dem in 2.1 beschriebenen Vorgehen von Welch (1939). Vergleichende Darstellungen für verschiedene nichtparametrische Dichteschätzungen geben Winter (1974), Goldstein (1975), Ambrosi (1976) und Victor (1976b). Am häufigsten wird eine *Kern-Methode* verwendet, d.h. es wird über jeden Stichprobenpunkt eine Dichtefunktion bekannter Form gelegt, der sogenannte *Kern*, und durch Mittelwertbildung über die Stichproben erhält man eine Schätzung der unbekannten Dichtefunktion. Der naheliegende Ansatz, nicht feste, sondern *variable Kerne* zu verwenden, scheint nach den Ergebnissen von Habbema, Hermans & Remme (1978) keine wesentliche Verbesserung zu bringen. Einen besonders einfachen Dichteschätzer, der auf Histogrammen beruht, untersuchen Chi & Ryzin (1977).

Als Spezialfall eines variablen Kernschätzers kann man das *Nächster-Nachbar-Verfahren* ansehen. Zu einem neu zu klassifizierenden Meßwertvektor zählt man aus, wie viele Meßwertvektoren aus den Ausgangsstichproben in einer Kugel um den zu klassifizierenden Vektor liegen. Je nach der Wahl des Distanzmaßes erhält man ein anderes Verfahren. Der Radius der Kugel kann der Lage des Meßwertvektors angepaßt werden, indem man den Radius z.B. so wählt, daß eine vorgegebene Anzahl von Stichprobenpunkten innerhalb jeder Kugel liegt. Da, wo die Punkte dicht liegen, ergibt sich dann ein kleiner Radius, in Randbereichen aber ein großer Radius. Nimmt man an, daß die Ausgangsstichproben alle den gleichen Umfang haben, so ordnet man das neu erhobene Individuum der Population zu, von der die meisten Stichprobenpunkte innerhalb der Kugel um dieses Individuum liegen. Von Goldstein (1972) werden die asymptotischen Eigenschaften dieses Verfahrens untersucht, während eine Simulationsstudie von Gessaman & Gessaman (1972) die Überlegenheit dieses Verfahrens im Vergleich zu mehreren anderen Verfahren zeigt.

Auf Dichteschätzungen beruhen auch die *Beste-Regel-der-Klasse-Verfahren* von Glick (1972). Hierbei wird unter allen Zuordnungsregeln einer vorgegebenen Klasse von Regeln diejenige bestimmt, für die die aus den Ausgangsstichproben geschätzte totale Wahrscheinlichkeit für eine richtige Zuordnung maximal ist.

Eine andere Klasse nichtparametrischer Zuordnungsregeln geht von Distanzmaßen zwischen empirischen Verteilungsfunktionen aus. Dabei liegt für jede

Population eine Ausgangsstichprobe vor, und eine neu erhobene Stichprobe soll einer der Populationen zugeordnet werden. Man berechnet für jede Stichprobe die empirische Verteilungsfunktion und ordnet die zu klassifizierende Stichprobe derjenigen Population zu, für die die Distanz der empirischen Verteilungsfunktionen minimal ist. Hierzu vergleiche man Hudimoto (1964) oder Das Gupta (1964).

Von Quesenberry & Gessaman (1968) wird vorgeschlagen, für jede Ausgangsstichprobe einen nichtparametrischen *Toleranzbereich* zu berechnen und neu erhobene Meßwertvektoren derjenigen Population zuzuordnen, in deren zugehörigen Toleranzbereich sie fallen. Meßwertvektoren, die in keinen Bereich oder in den Durchschnitt mehrerer Bereiche fallen, werden nicht klassifiziert. Falls man zuläßt, daß Individuen nicht klassifiziert werden, so spricht man von *partieller*, andernfalls von *erzwungener* Diskrimination. Von Anderson & Benning (1970) werden die Toleranzbereiche (für zwei Populationen) so bestimmt, daß die Wahrscheinlichkeit einer Nichtklassifikation minimiert wird, während Repges (1975) die Toleranzbereiche sequentiell bestimmt und dadurch eine sequentielle Diskrimination erhält.

Auch auf Rangzahlen beruhende Methoden werden in der Diskriminanzanalyse verwendet. So benutzen Fu & Chien (1967) sequentielle Ränge und Woinsky & Kurz (1969) die Differenz von Mann-Whitney-Statistiken zur sequentiellen Diskrimination. In Broffitt, Randles & Hogg (1976) und Randles, Broffitt, Ramberg & Hogg (1978a) werden Rangtransformationen angegeben, mit Hilfe derer man die üblichen linearen und quadratischen Diskriminanzfunktionen in verteilungsfreie Regeln transformieren kann. Dazu berechnet man für den zu klassifizierenden Meßwertvektor und jeden Vektor der Ausgangsstichproben die Diskriminanzfunktion und bestimmt für jede der beiden Stichproben von Funktionswerten den Rangplatz, den das zu klassifizierende Individuum einnimmt. Aufgrund dieser Ranginformationen wird das Individuum einer der beiden Populationen zugeordnet oder nicht klassifiziert. Jedoch kann man auch eine Version mit erzwungener Diskrimination vorsehen. Simulationsstudien für verschiedene Verteilungsannahmen zeigen, daß sich kleine Fehlerraten ergeben und eine Überlegenheit im Vergleich zu der Methode der Toleranzbereiche.

Weil die A-priori-Wahrscheinlichkeiten meist nicht bekannt sind, schätzt Bunke (1970) die unbekannten Dichten aus den empirischen Verteilungen und verwendet das Minimaxprinzip (vgl. 2.4). Die *Ordnungs-Statistik-Methode* von Kendall (1966) bestimmt zur Trennung von zwei Populationen  $P_1$  und  $P_2$  zunächst für eine Variable zwei Schranken. Wird die untere Schranke unterschritten, so wird ein Individuum  $P_1$  zugeordnet; bei Überschreitung der oberen Schranke wird es  $P_2$  zugeordnet. Für Werte im Zwischenbereich wird das Individuum noch nicht klassifiziert. Entsprechende Bedingungen werden auch

für die übrigen Variablen formuliert. Wenn viele Variablen gemessen werden, wird der Prozentsatz der bei diesem sequentiellen Verfahren nichtklassifizierten Individuen klein sein. Eine Verbesserung und Erweiterung des Verfahrens gibt Richards (1972) an.

Für die Trennung zweier Populationen bestimmen Fischer & Thiele (1979) die ungünstigsten Fehlklassifikationswahrscheinlichkeiten für lineare Diskriminanzfunktionen und minimieren diese dann. Speziell betrachten sie Neyman-Pearson-Regeln, für die bei festgehaltener Wahrscheinlichkeit für die eine Population die andere Wahrscheinlichkeit minimiert wird, sowie Minimax- und Bayes-Regeln. In Eye (1976) werden iterativ Trennebenen so bestimmt, daß die Anzahl der Fehlklassifikationen minimal wird.

Um eine Möglichkeit zu haben, unter der Vielzahl der nichtparametrischen Zuordnungsregeln eine geeignete auszuwählen, formulieren Fisher & Ness (1973) sieben Bedingungen, die man sinnvollerweise an Diskriminationsverfahren stellen kann, und geben für zehn verschiedene nichtparametrische Zuordnungsregeln jeweils an, ob für sie die Bedingungen erfüllt sind oder nicht.

## 5.2 Variablenauswahl

Für das Problem einer nichtparametrischen Variablenauswahl macht Whitney (1971) folgenden Ansatz. Zunächst schlägt er vor, die minimale Fehlerrate mit Hilfe der *Lasse-Einen-Aus-Methode* zu schätzen. Für zwei Populationen  $P_1$  bzw.  $P_2$  mit Ausgangsstichproben des Umfangs  $n_1$  bzw.  $n_2$  wählt man dazu zunächst ein Individuum aus der ersten Stichprobe aus und bestimmt mit Hilfe der restlichen ( $n_1 - 1 + n_2$ ) Individuen die Zuordnungsregel. Dieses macht man für alle  $n_1$  Individuen der ersten Stichprobe und wählt als Schätzung der Fehlklassifikationswahrscheinlichkeit die relative Häufigkeit der bei diesen  $n_1$  Klassifikationsversuchen fehlklassifizierten Individuen. Analog verfährt man für die zweite Stichprobe. Wie Lachenbruch & Mickey (1968) zeigen, ist dieses Verfahren der *Wiedereinsetzungsmethode*, bei der die Fehlerrate durch Einsetzen der Stichproben in die Diskriminanzfunktion, die mit Hilfe dieser Stichproben erhalten wurde, geschätzt wird, der *Aufteilungsmethode*, bei der die Diskriminanzfunktion mit Hilfe eines Teils der Individuen und die Fehlerrate mit Hilfe der restlichen Individuen geschätzt wird, sowie mehreren anderen Methoden überlegen. Whitney (1971) verwendet als nichtparametrische Zuordnungsregel ein *Nächster-Nachbar-Verfahren*. Um nicht alle Teilmengen von Variablen daraufhin überprüfen zu müssen, ob sich bei ihnen die geringste Schätzung der Fehlerrate ergibt, wird ein nichtoptimales Suchverfahren vorgeschlagen. Dazu wird zunächst die beste Einzelvariable ausgesucht, dann das beste Paar, das diese Einzelvariable enthält, usw.

### 5.3 Schätzungen der Fehlerrate

Sowohl bei der Auswahl eines nichtparametrischen Diskriminationsverfahrens, als auch bei der Variablenelektion ist es wichtig, über die Fehlerrate möglichst genaue Angaben zu haben. Besonders vorteilhaft ist die Kenntnis von nichtparametrischen Konfidenzintervallen für die Fehlerrate. Solche Konfidenzintervalle geben Devroye & Wagner (1976) für lineare Diskriminanzfunktionen für die Wiedereinsetzungsmethode an. Eine Erweiterung von linearen Zuordnungsregeln auf Nächster-Nachbar-Regeln und Histogramm-Regeln findet man bei Devroye & Wagner (1979b). Die Nächster-Nachbar-Regeln gehören zu den sogenannten *lokalen* Diskriminations-Regeln, weil jede Entscheidung nur von dem jeweiligen Individuum und seinen nächsten Nachbarn abhängt. Für solche Regeln leiten Rogers & Wagner (1978) und Devroye & Wagner (1979a) verteilungsfreie Konfidenzintervalle für die Fehlerrate her.

## 6. Analyse qualitativer und diskreter Daten

### 6.3 Verteilungsmodelle

#### 6.1.1 Volles Multinomialmodell

Sowohl die klassische Diskriminanzanalyse, die multivariate Normalverteilungen voraussetzt, als auch die meisten nichtparametrischen Verfahren verwenden wesentlich die Stetigkeit der zugrundeliegenden Meßwerte. Soweit es sich um diskrete Merkmale mit sehr vielen Ausprägungen handelt, mag die Verwendung dieser Verfahren approximativ gerechtfertigt sein. Jedoch liegen in der psychologischen Diagnostik, einem der häufigsten psychologischen Anwendungsbereiche für Diskriminanzanalysen, meist nur Merkmale mit sehr wenig Ausprägungen vor. Einem Testitem entsprechen meist weniger als zehn unterschiedliche Scores, so daß man von einem diskreten Merkmal ausgehen muß. Oft liegen auch nur Daten auf Nominalskalenniveau vor, so daß die Zuordnung von Testscores sehr willkürlich ist und die darin zum Ausdruck kommende Bewertung zu großen Fehlerraten Anlaß geben kann. Typische Beispiele sind Testitems mit den Antworten „Ja“ oder „Nein“, „Symptom vorhanden“ oder „Symptom nicht vorhanden“ usw. Einen Überblick über viele Diskriminationsverfahren für diskrete oder qualitative Daten gibt die Monographie von Goldstein & Dillen (1978).

Zunächst werde das *volle Multinomialmodell* für die Diskrimination von  $k$  Populationen  $P_1, \dots, P_k$  betrachtet. Die Variable  $X_1$  habe  $s_1$  Ausprägungen, die Variable  $X_2$  habe  $s_2$  Ausprägungen, usw. bis zur Variablen  $X_p$  mit  $s_p$  Ausprägungen. Insgesamt gibt es für die Meßwertvektoren  $s = s_1 \cdot s_2 \dots s_p$  verschiedene Ausprägungskombinationen oder *Zustände*, die mit den Wahrscheinlichkeiten

$p_1, \dots, p_s$  für die Population  $P_i$ ,  $i=1, \dots, k$  auftreten können. Falls man die Zustandswahrscheinlichkeiten  $p_{ij}$ ,  $i=1, \dots, k$ ,  $j=1, \dots, s$  und die A-priori-Wahrscheinlichkeiten  $q_i = P(P_i)$  kennt, ordnet man einen Meßwertvektor  $x' = x_1, \dots, x_p$ , der ja gleich einem der  $s$  Zustände  $j$  ist, derjenigen Population  $P_i$  zu, für die  $q_i p_{ij}$  maximal ist. Dieses entspricht der in 2.1 diskutierten Regel von Welch (1939). Sind die Wahrscheinlichkeiten nicht bekannt, so müssen sie aus den Ausgangsstichproben geschätzt werden. Falls man eine Zufallsstichprobe des Umfangs  $n$  aus der Gesamtpopulation zieht, von der  $n_i$  Individuen zu  $P_i$  gehören und  $n_{ij}$  Individuen, die zu  $P_i$  gehören, den Zustand  $j$  aufweisen, so schätzt man  $q_i$  durch  $n_i/n$  und  $p_{ij}$  durch  $n_{ij}/n_i$ . Der Diskriminanzscore  $q_i p_{ij}$  wird dann durch  $n_{ij}/n$  geschätzt. Von Glick (1973a) wird gezeigt, daß das beschriebene Verfahren eine tatsächliche Fehlerrate hat, die exponentiell gegen die optimale Fehlerrate konvergiert und daß die zu optimistische Verzerrung der scheinbaren Fehlerrate exponentiell gegen Null konvergiert. Ein Nachteil dieses vollen Multinomialmodells ist die Schwierigkeit, bei kleinen Ausgangsstichproben die Zustandswahrscheinlichkeiten reliabel zu schätzen, zumal wenn viele Zustände unbesetzt sind. Vor allem bei ungleichen A-priori-Wahrscheinlichkeiten oder ungleichen Stichprobenumfangen kann ein unbesetzter Zustand für eine Population etwas ganz anderes bedeuten als für eine andere Population. Der Einfluß kleiner Stichprobenumfänge wird von Cochran & Hopkins (1961) untersucht, die für nicht mehr als acht Zustände, also etwa für drei binäre Variablen, einen Stichprobenumfang von 50 für jede Population als ausreichend erachten.

### 6.1.2 Modelle bei multivariaten binären Items

Da beim vollen Multinomialmodell  $k \cdot (s-1)$  Zustandsparameter zu schätzen sind, nämlich für jede Population  $(s-1)$  unter Berücksichtigung der Tatsache, daß die Summe der Zustandswahrscheinlichkeiten jeweils 1 ist, versucht man das volle Modell weiter einzuschränken. Eine Möglichkeit ist die Beschränkung auf binäre Items, d.h. man setzt  $s_1 = \dots = s_k = 2$ . Dieses führt zu  $s = 2^p$  Zuständen oder  $(2^p - 1)$  Parametern für jede Population. Macht man zusätzlich die unrealistische Annahme, daß die Items unabhängig voneinander sind, so reduziert sich die Parameterzahl bei diesem *Interaktionsmodell 1. Ordnung* auf  $p$ . Für dieses Modell geben Goldstein & Dillon (1978) zwei Zuordnungsregeln an.

Von Bahadur (1961) wird ein Modell betrachtet, in dem explizit die Interaktionen höherer Ordnung auftreten. Durch Annahmen darüber, was für Interaktionen man noch berücksichtigen will, kann man in diesem **Bahadur-Modell** die Gesamtzahl der Parameter einschränken. Sind alle Interaktionen Null, so erhält man das Interaktionsmodell 1. Ordnung. Sind nur die Interaktionen 1. Ordnung von Null verschieden, so ergibt sich ein Interaktionsmodell

2. Ordnung, usw. Moore (1973b) vergleicht in einer Simulationsstudie das Verhalten des vollen Multinomialmodells, der Bahadur-Modelle 1. und 2. Ordnung sowie der linearen und quadratischen Diskriminanzfunktionen. Falls Interaktionen höherer Ordnung vorliegen, wird das volle Multinomialmodell empfohlen, sonst die lineare Diskriminanzfunktion.

Weit verbreitet ist das sogenannte *loglineare Modell*, bei dem die Logarithmen der Zustandswahrscheinlichkeiten als Linearkombination gewisser Haupteffekte und Interaktionen dargestellt werden. Wieder läßt sich die Anzahl der zu schätzenden Parameter reduzieren, wenn man annimmt, daß alle Wechselwirkungen ab einer gewissen Ordnung Null sind. Von Anderson (1972, 1974, 1979) werden Maximum-Likelihood-Schätzungen der Parameter angegeben.

Drei weitere Modelle, die die Zustandswahrscheinlichkeiten als Linearkombinationen von orthogonalen Polynomen darstellen, stammen von Martin & Bradley (1972), Ott & Kronmal (1976) und Goldstein (1977).

## 6.2 Nichtparametrische Verfahren bei qualitativen Daten

Von Aitchison & Aitken (1976) wird eine Art Kern definiert mit Hilfe eines Unähnlichkeitskoeffizienten zwischen den Meßwertvektoren. Dieses Verfahren wird verglichen mit dem Interaktionsmodell 1. Ordnung, mit der linearen Diskriminanzfunktion, mit dem loglinearen Modell, mit der logistischen Diskrimination und mit Nächster-Nachbar-Regeln. Weitere Vergleiche führt Aitken (1978) durch.

Dillon & Goldstein (1978) verwenden die Matusita-Distanz (Matusita, 1957) für diskrete Verteilungen. Diese bestimmt man, indem man zunächst die Wurzeln der Zustandswahrscheinlichkeiten für zwei Populationen berechnet, die Differenz dieser Wurzeln für jeweils gleiche Zustände bildet, diese Differenzen quadriert und aufsummiert. Diese Distanzen kann man auch berechnen, wenn man die Wahrscheinlichkeiten aus den Ausgangsstichproben schätzt. Man bestimmt die Distanz der beiden Ausgangsstichproben, wobei das zu klassifizierende Individuum einmal zu der ersten Stichprobe und einmal zu der zweiten Stichprobe gerechnet wird. Falls die erste Distanz größer ist, wird das Individuum der ersten Population zugeordnet, sonst der zweiten Population. Das Verfahren wird verglichen mit dem Martin-Bradley-Modell, der linearen Diskriminanzfunktion und dem Bahadur-Modell.

Von Lachin (1973) wird eine schrittweise Bayes-Regel vorgeschlagen, bei der die unbekannten Parameter nach der Maximum-Likelihood-Methode geschätzt werden. Der Bayes-Ansatz, d.h. die Klassifikation aufgrund der A-posteriori-Wahrscheinlichkeiten, wird auch von Victor, Trampisch & Zentgraf (1974) und Trampisch (1976, 1978) verwendet. Die Einschränkung des

vollen Multinomialmodells erfolgt hier durch das *Lancaster-Modell* unter Verwendung der Definition von Interaktionen höherer Ordnung nach Lancaster (1969). Von Hills (1967) wird ein Nächster-Nachbar-Verfahren zur nichtparametrischen Diskrimination diskreter Daten vorgeschlagen.

### 6.3 Gleichzeitiges Vorliegen diskreter und stetiger Variablen

Falls nicht nur Einzelitems, sondern auch Summenscores oder physiologische Variablen in einer Diskriminanzanalyse berücksichtigt werden sollen, sind die bisher beschriebenen Verfahren für qualitative Daten nicht anwendbar. Von Skene (1978) wird ein *latentes Klassen-Modell* im Sinne von Lazarsfeld & Henry (1968) zugrundegelegt, um die Parameterzahl zu reduzieren. Dieses Verfahren lässt sich auch dann anwenden, wenn die Meßwertvektoren der Ausgangsstichproben oder auch des zu klassifizierenden Individuums unvollständig sind. Die Parameter werden nach der Maximum-Likelihood-Methode mit Hilfe des EM-Algorithmus geschätzt. Dieses ist ein iteratives Verfahren, das auf jeder Stufe zuerst einen E-Schritt („expectation“) durchführt, bei dem Schätzungen der fehlenden Daten mit den bis dahin vorliegenden Parameterschätzungen erhalten werden. Im M-Schritt („maximisation“) werden dann aus dem vollständigen Datensatz neue Parameter geschätzt.

Press & Wilson (1978) verwenden für gleichzeitig vorliegende diskrete und stetige Variablen das Verfahren der logistischen Regression (vgl. 2.1, 2.9), während Krzanowski (1975, 1976) die Likelihood-Quotienten-Regel verwendet, wobei das volle Modell bei Vorliegen unbesetzter Zustände durch ein Interaktionsmodell 2. Ordnung approximiert wird. In Chang & Afifi (1974) wird ein Bayes-Ansatz vorgeschlagen.

### 6.4. Variablenauswahl

Auch bei qualitativen Daten ist es von Wichtigkeit, herauszufinden, welche Variablen am meisten zur Diskrimination beitragen und nur diese Variablen zu verwenden. Von Hills (1967) wird eine eingeschränkte schrittweise Vorwärtsauswahl empfohlen, wobei als Bewertungsmaß für die Diskriminationsstärke einer Variablenelementmenge das *Informationsdivergenzmaß* von Kuliback (1959) verwendet wird. Die Einschränkung besteht darin, daß in jedem Folgeschritt die Variablenmenge des vorangehenden Schritts weiterverwendet wird. Hills (1967) gibt auch eine nichteingeschränkte Prozedur an und diskutiert die Eigenschaften, die ein Bewertungsmaß für die Diskriminationsstärke haben sollte. Zwei solche Maße werden explizit angegeben. Eine Erweiterung dieser Verfahren stammt von Goldstein & Dillon (1977).

Von Glick (1973a) und Goldstein & Rabinowitz (1975) wird ein Bewertungsmaß, das mit der Matusita-Distanz (vgl. 6.2) in Zusammenhang steht, vorgeschlagen. Sowohl Raiffa (1961) als auch Lachin (1973) geben Vorwärtswahlverfahren an, die auf Bayes-Ansätze beruhen.

## Literatur

- Ahmann, J. S. 1955. An application of Fisher's discriminant function in the classification of students. *Journal of Educational Psychology*, 46, 184-188.
- Ahmed, S. W. & Lachenbruch, P. A. 1975. Discriminant analysis when one or both of the initial samples is contaminated: Large sample results. *EDV in Medizin und Biologie*, 6, 35-42.
- Ahmed, S. W. & Lachenbruch, P. A. 1977. Discriminant analysis when scale contamination is present in the initial Sample. In: J. van Ryzin (Hrsg.). *Classification and clustering*. 331-353. New York: Academic Press.
- Aitchison, J. & Aitken, C. G. G. 1976. Multivariate binary discrimination by the kernel method. *Biometrika*, 63, 413-420.
- Aitken, C. G. G. 1978. Methods of discrimination in multivariate binary data. In: L. C. A. Corsten & J. Hermans (Hrsg.). *COMPSTAT 1978*. 155-161. Wien: Physica.
- Alexakos, C. E. 1966. Predictive efficiency of two multivariate statistical techniques in comparison with clinical predictions. *Journal of Educational Psychology*, 57, 297-306.
- Altman, H., Evenson, R. C. & Cho, D. W. 1976. New discriminant functions for Computer diagnosis. *Multivariate Behavioral Research*, 11, 367-376.
- Ambrosi, K. 1976. Klassifikation mit Dichteschätzungen. *Operations Research Verfahren*, 22, 1-11.
- Anderson, H. E. 1966. Regression, discriminant analysis, and a Standard notation for basic statistics. In: R. B. Cattell (Hrsg.). *Handbook of multivariate experimental psychology*. 153-173. Chicago: Rand McNally.
- Andersort, J. A. 1969. Constrained discrimination between k populations. *Journal of the Royal Statistical Society, Series B* 31, 123-139.
- Anderson, J. A. 1972. Separate sample logistic discrimination. *Biometrika*, 59, 19-35.
- Anderson, J. A. 1973. Logistic discrimination with medical applications. In: T. Cacoullos (Hrsg.). *Discriminant analysis and applications*. 1-15. New York: Academic Press.
- Anderson, J. A. 1974. Diagnosis by logistic discriminant function: Further practical Problems and results. *Applied Statistics*, 23, 397-404.
- Anderson, J. A. 1975. Quadratic logistic discrimination. *Biometrika*, 62, 149-154.
- Anderson, J. A. 1979. Multivariate logistic compounds. *Biometrika*, 66, 17-26.

- Anderson, J. A. & Richardson, S. C. 1979. Logistic discrimination und bias correction in maximum likelihood estimation. *Technometrics*, 21, 71-78.
- Anderson, M. W. & Benning, R. D. 1970. A distribution-free discrimination procedure based on clustering. *IEEE Transactions on Information Theory* IT-16, 541-548.
- Anderson, T. W. 1951. Classification by multivariate analysis. *Psychometrika*, 16, 31-53.
- Anderson, T. W. 1958. An introduction to multivariate statistical analysis. New York: Wiley.
- Anderson, T. W. 1973a. Asymptotic evaluation of the probabilities of misclassification by linear discriminant functions. In: T. Cacoullos (Hrsg.). *Discriminant analysis and applications*. 17-35. New York: Academic Press.
- Anderson, T. W. 1973b. An asymptotic expansion of the distribution of the studentized classification statistic W. *Annals of Statistics*, 1, 964-972.
- Anderson, T. W. & Bahadur, R. R. 1962. Classification into two multivariate normal distributions with different covariance matrices. *Annals of Mathematical Statistics*, 33, 42-31.
- Aoyama, H. 1950. A note on the classification of observation data. *Annals of the Institute of Statistical Mathematics*, 2, 17-19.
- Armitage, P. 1950. Sequential analysis with more than two alternative hypotheses, and its relation to discriminant function analysis. *Journal of the Royal Statistical Society, Series B* 12, 137-144.
- Baggaley, A. R. & Campbell, J. P. 1967. Multiple-discriminant analysis of academic curricula by interest and aptitude variables. *Journal of Educational Measurement*, 4, 143-149.
- Bahadur, R. R. 1961. A representation of the joint distribution of responses to n dichotomous items. In: H. Solomon (Hrsg.). *Studies in item analysis and prediction*. 158-168. Stanford: University Press.
- Baitsch, H. & Bauer, R. K. 1956. Zum Problem der Merkmalsauswahl für Trennverfahren (Barnard-Problem). *Allgemeines Statistisches Archiv*, 40, 160-167.
- Balakrishnan, V. & Sanghvi, L. D. 1968. Distance between populations on the basis of attribute data. *Biometrics*, 24, 859-865.
- Bandyopadhyay, S. 1979. Two population classification in Gaussian process. *Journal of Statistical Planning and Inference*, 3, 225-233.
- Banerjee, K. S. & Marcus, L. F. 1965. Bounds in a minimax classification procedure. *Biometrika*, 52, 653-654.
- Bartlett, M. S. & Please, N. W. 1963. Discrimination in the case of zero mean differences. *Biometrika*, 50, 17-21.
- Bauer, R. K. 1954. Diskriminanzanalyse. *Allgemeines Statistisches Archiv*, 38, 205-216.
- Beall, G. 1945. Approximate methods in calculating discriminant functions. *Psychometrika*, 10, 205-217.

- Behrens, W. U. 1959. Beitrag zur Diskriminanzanalyse. *Biometrische Zeitschrift*, 1, 3-14.
- Ben-Bassat, M. & Gal, S. 1977. Properties and convergence of a posteriori probabilities in classification Problems. *Pattern Recognition*, 9, 99-107.
- Bhattacharya, P. K. & Das Gupta, S. 1964. Classification between univariate exponential populations. *Sankhyā, Series A* 26, 17-24.
- Bledsoe, J. C. 1973. The prediction of teacher competence: A comparison of two multivariate statistical techniques. *Multivariate Behavioral Research*, 8, 3-22.
- Borgen, F. H. & Seling, M. J. 1978. Uses of discriminant analysis following MANOVA: Multivariate statistics for multivariate purposes. *Journal of Applied Psychology*, 63, 689-697.
- Bowker, A. H. 1961. A representation of Hotelling's  $T^2$  and Anderson's classification statistic W in terms of simple statistics. In: H. Solomon (Hrsg.). *Studies in item analysis and prediction*. 285-292. Stanford: University Press.
- Bowker, A. H. & Sitgreaves, R. 1961. An asymptotic expansion for the distribution function of the W-classification statistic. In: H. Solomon (Hrsg.). *Studies in item analysis and prediction*. 293-310. Stanford: University Press.
- Broffitt, J. D., Randles, R. H. & Hogg, R. V. 1976. Distribution-free partial discriminant analysis. *Journal of the American Statistical Association*, 71, 934-939.
- Brown, G. W. 1947. Discriminant functions. *Annals of Mathematical Statistics*, 18, 514-528.
- Brown, G. W. 1950. Basic principles for construction and application of discriminators. *Journal of Clinical Psychology*, 6, 58-61.
- Bryan, J. G. 1951. The generalized discriminant function: Mathematical foundation and computational routine. *Harvard Educational Review*, 21, 90-95.
- Bunke, O. 1967. Stabilität statistischer Entscheidungsprobleme und Anwendungen in der Diskriminanzanalyse. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 7, 131-146.
- Bunke, O. 1970. Non-parametric decision functions with asymptotic minimax optimality. *Theory of Probability and its Applications*, 15, 148-153.
- Burnaby, T. P. 1966. Growth-invariant discriminant functions and generalized distances. *Biometrics*, 22, 96-110.
- Cacoullos, T. (Hrsg.). 1973a. *Discriminant analysis and applications*. New York: Academic Press.
- Cacoullos, T. 1973b. Distance, discrimination and error. In: T. Cacoullos (Hrsg.). *Discriminant analysis and applications*. 61-75. New York: Academic Press.
- Cacoullos, T. & Styan, G. P. H. 1973. A bibliography of discriminant analysis. In: T. Cacoullos (Hrsg.). *Discriminant analysis and applications*. 375-434. New York: Academic Press.
- Chan, L. S. & Dunn, O. J. 1972. The treatment of missing values in discriminant analysis - I. The sampling experiment. *Journal of the American Statistical Association*, 67, 473-477.

- Chang, P. C. & Afifi, A. A. 1974. Classification based on dichotomous and continuous variables. *Journal of the American Statistical Association*, 69, 336-339.
- Chi, P. Y. & Ryzin, J. van. 1977. A simple histogram method for nonparametric classification. In: J. van Ryzin (Hrsg.). *Classification and clustering*. 395-421. New York: Academic Press.
- Chikuse, Y. 1978. Asymptotic distributions of the latent roots with multiple population roots in multiple discriminant analysis. *Annals of the Institute of Statistical Mathematics*, Part A 29, 57-62.
- Chinganda, E. F. & Subrahmaniam, K. 1979. Robustness of the linear discriminant function to nonnormality: Johnson's System. *Journal of Statistical Planning and Inference*, 3, 69-77.
- Chou, R. J. & Muirhead, R. J. 1979. On some distribution problems in MANOVA and discriminant analysis. *Journal of Multivariate Analysis*, 9, 410-419.
- Clarke, W. R., Lachenbruch, P. A. & Broffitt, B. 1979. How non-normality affects the quadratic discriminant function. *Communications in Statistics - Theory and Methods* A8, 1285-1301.
- Clavelle, P. R. & Butcher, J. N. 1977. An adaptive typological approach to psychiatric screening. *Journal of Consulting and Clinical Psychology*, 45, 851-859.
- Cleveland, W. S. & Lachenbruch, P. A. 1974. A measure of divergence among several populations. *Communications in Statistics*, 3, 201-211.
- Clunies-Ross, C. W. & Riffenburgh, R. H. 1960. Geometry and linear discrimination. *Biometrika*, 47, 185-189.
- Cochran, W. G. 1964 a. On the Performance of the linear discriminant function. *Tech-nometrics*, 6, 179-190.
- Cochran, W. G. 1964 b. Comparison of two methods of handling covariates in discriminatory analysis. *Annals of the Institute of Statistical Mathematics*, 16, 43-53.
- Cochran, W. G. & Bliss, C. I. 1948. Discriminant functions with covariance. *Annals of Mathematical Statistics*, 19, 151-176.
- Cochran, W. G. & Hopkins, C. E. 1961. Some classification problems with multivariate qualitative data. *Biometrics*, 17, 10-32.
- Coolidge, F. L. 1976. Discriminant and factor analysis of the WAIS and the Satz-Mogel abbreviated WAIS on brain-damaged and psychiatric patients. *Journal of Consulting and Clinical Psychology*, 44, 153.
- Cooper, P. W. 1963. Statistical classification with quadratic forms. *Biometrika*, 50, 439-448.
- Cox, D. R. 1966. Some procedures connected with the logistic qualitative response curve. In: F. N. David (Hrsg.). *Research papers in statistics*. 55-71. London: Wiley.
- Cramer, E. M. 1967. Equivalence of two methods of computing discriminant function coefficients. *Biometrics*, 23, 153.

- Cramer, E. M. 1975. The relation between Rao's paradox in discriminant analysis and regression analysis. *Multivariate Behavioral Research*, 10, 99-107.
- Das Gupta, S. 1964. Non-parametric classification rules. *Sankhyā, Series A* 26, 25-30.
- Das Gupta, S. 1965. Optimum classification rules for classification into two multivariate normal populations. *Annals of Mathematical Statistics*, 36, 1174-1184.
- Das Gupta, S. 1968. Some aspects of discrimination function coefficients. *Sankhyā, Series A* 30, 387-400.
- Das Gupta, S. 1973. Theories and methods in classification: A review. In: T. Cacoullos (Hrsg.). *Discriminant analysis and applications*. 77-137. New York: Academic Press.
- Day, N. E. 1969. Linear and quadratic discrimination in pattern recognition. *IEEE Transactions on Information Theory* IT-15, 419-420.
- Day, N. E. & Kerridge, D. F. 1967. A general maximum likelihood discriminant. *Biometrics*, 23, 313-323.
- Desu, M. M. & Geisser, S. 1973. Methods and applications of equal-mean discrimination. In: T. Cacoullos (Hrsg.). *Discriminant analysis and applications*. 139-159. New York: Academic Press.
- Devroye, L. P. & Wagner, T. J. 1976. A distribution-free performance bound in error estimation. *IEEE Transactions on Information Theory* IT-22, 586-590.
- Devroye, L. P. & Wagner, T. J. 1979a. Distribution-free inequalities for the deleted and holdout error estimates. *IEEE Transactions on Information Theory* IT-25, 202-207.
- Devroye, L. P. & Wagner, T. J. 1979b. Distribution-free performance bounds with the resubstitution error estimate. *IEEE Transactions on Information Theory* IT-25, 208-210.
- Dillon, W. R. & Goldstein, M. 1978. On the Performance of some multinomial classification rules. *Journal of the American Statistical Association*, 73, 305-313.
- DiPillo, P. J. 1976. The application of bias to discriminant analysis. *Communications in Statistics - Theory and Methods* A5, 843-854.
- Dixon, W. J. (Hrsg.). 1976. *Biomedical computer programs-BMD*. Berkeley: University of California Press.
- Dudzinski, M. L. 1977. The detection of adolescent behaviour in rabbits. *Mathematical Scientist*, 2, 39-47.
- Dunn, O. J. & Varady, P. D. 1966. Probabilities of correct classification in discriminant analysis. *Biometrics*, 22, 908-924.
- Dunteman, G. H. 1966. Discriminant analyses of the SVIB for female students in five college curricula. *Journal of Applied Psychology*, 50, 509-515.
- Efron, B. 1975. The efficiency of logistic regression compared to normal discriminant analysis. *Journal of the American Statistical Association*, 70, 892-898.
- Eisenbeis, R. A., Gilbert, G. G. & Avery, R. B. 1973. Investigating the relative importance of individual variables and variable subsets in discriminant analysis. *Communications in Statistics*, 2, 205-219.

- Elfving, G. 1961. An expansion principle for distribution functions with applications to Student's statistic and the one-dimensional classification statistic. In: H. Solomon (Hrsg.). *Studies in item analysis and prediction*. 276-284. Stanford: University Press.
- Ellison, B. E. 1962. A classification problem in which information about alternative distributions is based on samples. *Annals of Mathematical Statistics*, 33, 213-223.
- Ellison, B. E. 1965. Multivariate classification with covariances known. *Annals of Mathematical Statistics*, 36, 1787-1793.
- Enis, P. & Geisser, S. 1971. Optimal predictive linear discriminants. Abstract. *Annals of Mathematical Statistics*, 42, 2179.
- Enis, P. & Geisser, S. 1974. Optimal predictive linear discrimination. *Annals of Statistics*, 2, 403-410.
- Eye, A. von 1976. Ein iteratives Verfahren zur Lösung des linearen Diskriminanzproblems. *Psychologische Beiträge*, 18, 190-207.
- Eye, A. von & Hüssy, W. 1979. Zum Beitrag von Variablen der kognitiven Komplexität für die Identifikation verkehrspychologischer Risikogruppen. *Schweizerische Zeitschrift für Psychologie*, 38, 58-70.
- Eye, A. von & Wiedl, K. H. 1978. Personentypen ästhetischer Präferenz und ihre Klassifikationseigenschaften. *Zeitschrift für experimentelle und angewandte Psychologie*, 25, 349-366.
- Eysenck, H. J. 1955. Psychiatric diagnosis as a psychological and statistical problem. *Psychological Reports*, 1, 3-17.
- Eysenck, H. J. & Claridge, G. 1962. The position of hysterics and dysthymics in a two-dimensional framework of personality description. *Journal of Abnormal and Social Psychology*, 64, 46-55.
- Farver, T. B. & Dunn, O. J. 1979. Stepwise variable selection in classification problems. *Biometrical Journal*, 21, 145-153.
- Fischer, K. & Thiele, C. 1979. On a distributionfree method in discriminant analysis. *Mathematische Operationsforschung und Statistik, Series Statistics*, 10, 281-289.
- Fisher, L. & Ness, J. W. van 1973. Admissible discriminant analysis. *Journal of the American Statistical Association*, 68, 603-607.
- Fisher, R. A. 1936. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7, 179-188.
- Fisher, R. A. 1938. The statistical utilization of multiple measurements. *Annals of Eugenics*, 8, 376-386.
- Freedman, D. A. 1967. A remark on sequential discrimination. *Annals of Mathematical Statistics*, 38, 1666-1670.
- Fu, K. S. & Chien, Y. T. 1967. Sequential recognition using a nonparametric ranking procedure. *IEEE Transactions on Information Theory* IT-13, 484-492.
- Ganesalingam, S. & McLachlan, G. J. 1978. The efficiency of a linear discriminant function based on unclassified initial samples. *Biometrika*, 65, 658-662.

- Ganesalingam, S. & McLachlan, G. J. 1979. Small sample results for a linear discriminant function estimated from a mixture of normal populations. *Journal of Statistical Computation and Simulation*, 9, 151-158.
- Garrett, H. E. 1943. The discriminant function and its use in psychology. *Psychometrika*, 8, 65-79.
- Geisser, S. 1964. Posterior odds for multivariate normal classifications. *Journal of the Royal Statistical Society, Series B* 26, 69-76.
- Geisser, S. 1966. Predictive discrimination. In: P. R. Krishnaiah (Hrsg.). *Multivariate analysis*. 149-163. New York: Academic Press.
- Geisser, S. 1973. A note on linear discriminants. *Bulletin of the International Statistical Institute* 45 (I), 442-448.
- Geisser, S. 1977. Discrimination, allocatory and separatory, linear aspects. In: J. van Ryzin (Hrsg.). *Classification and clustering*. 301-330. New York: Academic Press.
- Geisser, S. & Desu, M. M. 1968. Predictive zero-mean uniform discrimination. *Biometrika*, 55, 519-524.
- Gessaman, M. P. & Gessaman, P. H. 1972. A comparison of some multivariate discriminant procedures. *Journal of the American Statistical Association*, 67, 468-472.
- Glick, N. 1972. Sample-based classification procedures derived from density estimators. *Journal of the American Statistical Association*, 67, 116-122.
- Glick, N. 1973 a. Sample-based multinomial classification. *Biometrics*, 29, 241-256.
- Glick, N. 1973b. Separation and probability of correct classification among two or more distributions. *Annals of the Institute of Statistical Mathematics*, 25, 373-382.
- Goldman, R. D. & Warren, R. 1973. Discriminant analysis of study strategies connected with college grade success in different major fields. *Journal of Educational Measurement*, 10, 39-47.
- Goldstein, M. 1972.  $k_n$ -nearest neighbor classification. *IEEE Transactions on Information Theory* IT-18, 627-630.
- Goldstein, M. 1975. Comparison of some density estimate classification procedures. *Journal of the American Statistical Association*, 70, 666-669.
- Goldstein, M. 1976. An approximate test for comparative discriminatory power. *Multivariate Behavioral Research*, 11, 157-163.
- Goldstein, M. 1977. A two-group classification procedure for multivariate dichotomous responses. *Multivariate Behavioral Research*, 12, 335-346.
- Goldstein, M. & Dillen, W. R. 1977. A stepwise discrete variable selection procedure. *Communications in Statistics - Theory and Methods* A6, 1423-1436.
- Goldstein, M. & Dillen, W. R. 1978. *Discrete discriminant analysis*. New York: Wiley.
- Goldstein, M. & Rabinowitz, M. 1975. Selection of variates for the two-group multinomial classification problem. *Journal of the American Statistical Association*, 70, 776-781.

- Gupta, A. K. 1977. On the equivalence of two classification rules. *Biometrical Journal*, 19, 365-367.
- Gupta, S. S. & Huang, D. Y. 1975. On  $\Gamma$ -minimax classification procedures. *Bulletin of the International Statistical Institute* 46 (3), 330-335.
- Guseman, L. F., Peters, B. C. & Walker, H. F. 1975. On minimizing the probability of misclassification for linear feature selection. *Annals of Statistics*, 3, 661-668.
- Habbema, J. D. F. & Hermans, J. 1977. Selection of variables in discriminant analysis by F-statistic and error rate. *Technometrics*, 19, 487-493.
- Habbema, J. D. F., Hermans, J. & Broek, K. van den 1974. A stepwise discriminant analysis program using density estimation. In: G. Bruckmann, F. Ferschl & L. Schmetterer (Hrsg.). COMPSTAT 1974. 101-110. Wien: Physica.
- Habbema, J. D. F., Hermans, J. & Remme, J. 1978. Variable kernel density estimation in discriminant analysis. In: L. C. A. Corsten & J. Hermans (Hrsg.). COMPSTAT 1978. 178-185. Wien: Physica.
- Han, C. P. 1968. A note on discrimination in the case of unequal covariance matrices. *Biometrika*, 55, 586-587.
- Han, C. P. 1969. Distribution of discriminant function when covariance matrices are proportional. *Annals of Mathematical Statistics* 40, 979-985.
- Han, C. P. 1970. Distribution of discriminant function in circular models. *Annals of the Institute of Statistical Mathematics*, 22, 117-125.
- Hand, H. H. & LaFollette, W. R. 1973. A discriminant analysis of organizational Performance variables. *Journal of Applied Psychology*, 58, 368-371.
- Harper, A. E. 1950a. Discrimination of the types of schizophrenia by the Wechsler-Bellevue scale. *Journal of Consulting Psychology*, 14, 290-296.
- Harper, A. E. 1950 b. Discrimination between matched schizophrenics and normals by the Wechsler-Bellevue scale. *Journal of Consulting Psychology*, 14, 351-357.
- Harter, H. L. 1951. On the distribution of Wald's classification statistic. *Annals of Mathematical Statistics*, 22, 58-67.
- Hattemer, H. 1974. Ein Zusammenhang zwischen Diskriminanz- und Regressionsanalyse. *EDV in Medizin und Biologie*, 5, 7-10.
- Healy, M. J. R. 1965. Computing a discriminant function from within-sample dispersions. *Biometrics*, 21, 1011-1012.
- Hecker, R. & Wegener, H. 1978. The valuation of classification rates in stepwise discriminant analyses. *Biometrical Journal*, 20, 713-727.
- Henschke, C. I. & Chen, M. M. 1974. Variable selection technique for classification Problems. *Educational and Psychological Measurement*, 34, 11-18.
- Hermans, J. & Habbema, J. D. F. 1975. Comparison of five methods to estimate posterior probabilities. *EDV in Medizin und Biologie*, 6, 1919.
- Hildebrandt, B., Michaelis, J. & Koller, S. 1973. Die Häufigkeit der Fehlklassifikation bei der quadratischen Diskriminanzanalyse. *Biometrische Zeitschrift*, 15, 3-12.
- Hills, M. 1966. Allocation rules and their error rates. *Journal of the Royal Statistical Society, Series B* 28, 1-20.

- Hills, M. 1967. Discrimination and allocation with discrete data. *Applied Statistics*, 16, 237-250.
- Hoel, P. G. & Peterson, R. P. 1949. A solution to the problem of Optimum classification. *Annals of Mathematical Statistics*, 20, 433-438.
- Hora, S. C. 1978. Sample size determination in Bayesian discriminant analysis. *Journal of the American Statistical Association*, 73, 569-572.
- Horst, P. 1956a. Multiple classification by the method of least squares. *Journal of Clinical Psychology*, 12, 3-16.
- Horst, P. 1956b. Least square multiple classification for unequal subgroups. *Journal of Clinical Psychology*, 12, 309-315.
- Horst, P. & Smith, S. 1950. The discrimination of two racial samples. *Psychometrika*, 15, 271-289.
- Huberty, C. J. 1975. Discriminant analysis. *Review of Educational Research*, 45, 543-598.
- Huberty, C. J. & Curry, A. R. 1978. Linear versus quadratic multivariate classification. *Multivariate Behavioral Research*, 13, 237-245.
- Hudimoto, H. 1964. On a distribution-free two-way classification. *Annals of the Institute of Statistical Mathematics*, 16, 247-253.
- Hudlet, R. & Johnson, R. 1977. Linear discrimination and some further results on best lower dimensional representations. In: J. van Ryzin (Hrsg.). *Classification and Clustering*. 371-393. New York: Academic Press.
- Isaacson, S. L. 1964. Problems in classifying populations. In: O. Kempthorne, T. A. Bancroft, J. W. Gowen & J. L. Lush (Hrsg.). *Statistics and Mathematics in Biology*. 107-117. New York: Hafner.
- Isii, K. & Taga, Y. 1978. Bound on the classification error for discriminating between multivariate populations with specified means and covariance matrices. *Annals of Statistics*, 6, 132-141.
- Jackson, R. 1950. The selection of students for freshman chemistry by means of discriminant functions. *Journal of Experimental Education*, 18, 209-214.
- Janke, W. 1971. Klassifikation. In: R. Heiss (Hrsg.). *Handbuch der Psychologie*. Band 6. *Psychologische Diagnostik*. 901-929. Göttingen: Hogrefe.
- John, S. 1959. The distribution of Wald's classification statistic when the dispersion matrix is known. *Sankhyä*, 21, 371-376.
- John, S. 1960a. On some classification problems - I. *Sankhyä*, 22, 301-308.
- John, S. 1960b. On some classification statistics. *Sankhyä*, 22, 309-316.
- John, S. 1961. Errors in discrimination. *Annals of Mathematical Statistics*, 32, 1125-1144.
- John, S. 1962-1963. On classification by the statistics R and Z. *Annals of the Institute of Statistical Mathematics*, 14, 237-246.
- John, S. 1964. Further results on classification by W. *Sankhyä, Series A* 26, 39-46.

- Johnson, M. C. 1955. Classification by multivariate analysis with objectives of minimizing risk, minimizing maximum risk, and minimizing probability of misclassification. *Journal of Experimental Education*, 23, 259-264.
- Kashyap, R. L. 1978. Optimal feature selection and decision rules in classification problems with time series. *IEEE Transactions on Information Theory* IT-24, 281-288.
- Kelly, F. J., Veldman, D. J. & McGuire, C. 1964. Multiple discriminant prediction of delinquency and school dropouts. *Educational and Psychological Measurement*, 24, 535-544.
- Kendall, M. G. 1966. Discrimination and classification. In: P. R. Krishnaiah (Hrsg.). *Multivariate analysis*. 165-185. New York: Academic Press.
- Kendall, M. G. & Stuart, A. 1966. The advanced theory of statistics. Volume 3. Design and analysis, and time-series. London: Griffin.
- Koffler, S. L. & Penfield, D. A. 1979. Nonparametric discrimination procedures for non-normal distributions. *Journal of Statistical Computation and Simulation*, 8, 281-299.
- Kossack, C. F. 1945. On the mechanics of classification. *Annals of Mathematical Statistics*, 16, 95-98.
- Kossack, C. F. 1963. Statistical classification techniques. *IBM Systems Journal*, 2, 136-151.
- Krzanowski, W. J. 1975. Discrimination and classification using both binary and continuous variables. *Journal of the American Statistical Association*, 70, 782-790.
- Krzanowski, W. J. 1976. Canonical representation of the location model for discrimination or classification. *Journal of the American Statistical Association*, 71, 845-848.
- Krzanowski, W. J. 1977. The Performance of Fisher's linear discriminant function under non-optimal conditions. *Technometrics*, 19, 191-200.
- Kshirsagar, A. M. 1970. Distributions associated with the factors of Wilks' lambda in discriminant analysis. *Journal of the Australian Mathematical Society*, 10, 269-277.
- Kshirsagar, A. M. 1971. Goodness of fit of a discriminant function from the vector space of dummy variables. *Journal of the Royal Statistical Society, Series B* 33, 111-116.
- Kshirsagar, A. M. & Arseven, E. 1975. A note on the equivalence of two discrimination procedures. *American Statistician*, 29, 38-39.
- Kudo, A. 1959. The classificatory problem viewed as a two-decision problem. *Memoirs of the Faculty of Science, Kyushu University, Series A*, 13, 96-125.
- Kuliback, S. 1959. *Information theory and statistics*. New York: Wiley.
- Lachenbruch, P. A. 1966. Discriminant analysis when the initial samples are misclassified. *Technometrics*, 8, 657-662.

- Lachenbruch, P. A. 1968. On expected probabilities of misclassification in discriminant analysis, necessary sample size, and a relation with the multiple correlation coefficient. *Biometrics*, 24, 823-834.
- Lachenbruch, P. A. 1973. Some results on the multiple group discriminant problem. In: T. Cacoullos (Hrsg.). *Discriminant analysis and applications*. 193-205. New York: Academic Press.
- Lachenbruch, P. A. 1974. Discriminant analysis when the initial samples are misclassified II: Non-random misclassification models. *Technometrics*, 16, 419-424.
- Lachenbruch, P. A. 1975. *Discriminant analysis*. New York: Hafner.
- Lachenbruch, P. A. 1977. Covariance adjusted discriminant functions. *Annals of the Institute of Statistical Mathematics*, Part A 29, 247-257.
- Lachenbruch, P. A. 1979. Note on initial misclassification effects on the quadratic discriminant function. *Technometrics*, 21, 129-132.
- Lachenbruch, P. A. & Goldstein, M. 1979. Discriminant analysis. *Biometrics*, 35, 69-85.
- Lachenbruch, P. A. & Kupper, L. L. 1973. Discriminant analysis when one population is a mixture of normals. *Biometrische Zeitschrift*, 15, 191-197.
- Lachenbruch, P. A. & Mickey, M. R. 1968. Estimation of error rates in discriminant analysis. *Technometrics*, 10, 1-11.
- Lachirr, J. M. 1973. On a stepwise procedure for two population Bayes decision rules using discrete variables. *Biometrics*, 29, 551-564.
- Lachin, J. M. & Schachter, J. 1974. On stepwise discriminant analyses applied to physiological data. *Psychophysiology*, 11, 703-709.
- Ladd, G. W. 1966. Linear probability functions and discriminant functions. *Econometrica*, 34, 873-885.
- Lancaster, H. O. 1969. *The chi-squared distribution*. New York: Wiley.
- Lazarsfeld, P. F. & Henry, N. W. 1968. *Latent structure analysis*. Boston: Houghton Mifflin.
- Leton, D. A. & Anderson, H. E. 1964. Discriminant analysis of achievement characteristics for multi-grade grouping of students. *Journal of Experimental Education*, 32, 293-297.
- Linhart, H. 1961. Zur Wahl von Variablen in der Trennanalyse. *Metrika*, 4, 126-139.
- Lissitz, R. W. & Henschke-Mason, C. 1972. The selection of independent variables and prior probabilities as a factor influencing the accuracy of classifying individuals to existing groups. *Multivariate Behavioral Research*, 7, 489-497.
- Lohnes, P. R. 1961. Test space and discriminant space classification models and related significance tests. *Educational and Psychological Measurement*, 21, 559-574.
- MacFadyen, H. W. 1975. The classification of depressive disorders. *Journal of Clinical Psychology*, 31, 380-401.
- Malgady, R. G. 1977. Discriminant analysis of psychological judgements of literal and figurative meaningfulness and anomaly. *Journal of Psychology*, 95, 217-221.

- Mallinger, B. L. 1977. Background inference procedure and discriminant function analysis in predicting clinically determined categories of learning disability. *Perceptual and Motor Skills*, 44, 767-776.
- Mallows, C. L. 1953. Sequential discrimination. *Sankhyä*, 12, 321-338.
- Marks, S. & Dunn, O. J. 1974. Discriminant functions when covariance matrices are unequal. *Journal of the American Statistical Association*, 69, 555-559.
- Marshall, A. W. & Olkin, I. 1968. A general approach to some screening and classification problems. *Journal of the Royal Statistical Society, Series B* 30, 407-443.
- Martin, D. C. & Bradley, R. A. 1972. Probability models, estimation, and classification for multivariate dichotomous populations. *Biometrics*, 28, 203-221.
- Matusita, K. 1957. Decision rule, based on the distance, for the classification problem. *Annals of the Institute of Statistical Mathematics*, 8, 67-77.
- Matusita, K. 1964. Distance and decision rules. *Annals of the Institute of Statistical Mathematics*, 16, 305-315.
- Matusita, K. 1967. Classification based on distance in multivariate Gaussian cases. In: L. M. Le Cam & J. Neyman (Hrsg.). *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Volume I. 299-304. Berkeley: University of California Press.
- Matusita, K. 1971. Some properties of affinity and applications. *Annals of the Institute of Statistical Mathematics*, 23, 137-155.
- Matusita, K. 1973. Discrimination and the affinity of distributions. In: T. Cacoullos (Hrsg.). *Discriminant analysis and applications*. 213-223. New York: Academic Press.
- McCabe, G. P. 1975. Computations for variable selection in discriminant analysis. *Technometrics*, 17, 103-109.
- McDonald, L. L., Lowe, V. W., Smidt, R. K. & Meister, K. A. 1976. A preliminary test for discriminant analysis based on small samples. *Biometrics*, 32, 417-422.
- McKay, R. J. 1976. Simultaneous procedures in discriminant analysis involving two groups. *Technometrics*, 18, 47-53.
- McKay, R. J. 1977. Simultaneous procedures for variable selection in multiple discriminant analysis. *Biometrika*, 64, 283-290.
- McKay, R. J. 1978. A graphical aid to selection of variables in two-group discriminant analysis. *Applied Statistics*, 27, 259-263.
- McKeachie, W. J. & Lin, Y. G. 1975. Multiple discriminant analysis of student ratings of college teachers. *Journal of Educational Research*, 68, 300-305.
- McLachlan, G. J. 1972. An asymptotic expansion for the variance of the errors of misclassification of the linear discriminant function. *Australian Journal of Statistics*, 14, 68-72.
- McLachlan, G. J. 1974. The asymptotic distributions of the conditional error rate and risk in discriminant analysis. *Biometrika*, 61, 131-135.

- McLachlan, G. J. 1975a. Some expected values for the error rates of the sample quadratic discriminant function. *Australian Journal of Statistics*, 17, 161-165.
- McLachlan, G. J. 1975b. Iterative reclassification procedure for constructing an asymptotically optimal rule of allocation in discriminant analysis. *Journal of the American Statistical Association*, 70, 365-369.
- McLachlan, G. J. 1976. A criterion for selecting variables for the linear discriminant function. *Biometrics*, 32, 529-534.
- McLachlan, G. J. 1977a. Estimating the linear discriminant function from initial samples containing a small number of unclassified observations. *Journal of the American Statistical Association*, 72, 403-406.
- McLachlan, G. J. 1977b. Constrained sample discrimination with the studentized classification statistic W. *Communications in Statistics - Theory and Methods* A6, 575-583.
- McLachlan, G. J. 1978. Small sample results for partial classification with the studentized statistic W. *Biometrical Journal*, 20, 639-644.
- Melrose, J. P., Stroebel, C. F. & Glueck, B. C. 1970. Diagnosis of psychopathology using stepwise multiple discriminant analysis. 1. *Comprehensive Psychiatry*, 11, 43-50.
- Memon, A. Z. & Okamoto, M. 1970. The classification statistic  $W^*$  in covariate discriminant analysis. *Annals of Mathematical Statistics*, 41, 1491-1499.
- Memon, A. Z. & Okamoto, M. 1971. Asymptotic expansion of the distribution of the Z statistic in discriminant analysis. *Journal of Multivariate Analysis*, 1, 294-307.
- Michaelis, J. 1973a. Simulation experiments with multiple group linear and quadratic discriminant analysis. In: T. Cacoullos (Hrsg.). *Discriminant analysis and applications*. 225-238. New York: Academic Press.
- Michaelis, J. 1973b. Beurteilung von Trennergebnissen bei praktischen Anwendungen der Diskriminanzanalyse. In: H. J. Lange & G. Wagner (Hrsg.). *Computergestützte ärztliche Diagnostik*. 261-267. Stuttgart: Schattauer.
- Mises, R. von. 1945. On the classification of Observation data into distinct groups. *Annals of Mathematical Statistics*, 16, 68-73.
- Moore, D. H. 1973a. Combining linear and quadratic discriminants. *Computers and Biomedical Research*, 6, 422-429.
- Moore, D. H. 1973b. Evaluation of five discrimination procedures for binary variables. *Journal of the American Statistical Association*, 68, 399-404.
- Moran, M. A. 1975a. The effects of selecting variables for use in the linear discriminant function. *EDV in Medizin und Biologie*, 6, 24-30.
- Moran, M. A. 1975b. On the expectation of errors of allocation associated with a linear discriminant function. *Biometrika*, 62, 141-148.
- Nakache, J. P. & Dusserre, L. 1975. Practical problems in linear discriminant analysis. *EDV in Medizin und Biologie*, 6, 30-35.
- Narain, R. D. 1950. Some results on discriminant functions. *Journal of the Indian Society of Agricultural Statistics*, 2, 49-59.

- Neymark, Y. I., Breydo, M. D. & Durnovo, A. N. 1970. A linear minimax classification algorithm. *Engineering Cybernetics*, Nr. 2, 328-336.
- Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K. & Bent, D. H. 1975. Statistical package for the social sciences. New York: McGraw-Hill.
- Nishi, A. 1977. On linear classification procedures between two categories with known mean vectors and covariance matrices. *Annals of the Institute of Statistical Mathematics*, Part A 29, 433-444.
- Okamoto, M. 1963. An asymptotic expansion for the distribution of the linear discriminant function. *Annals of Mathematical Statistics*, 34, 1286-1301.
- O'Neill, T. J. 1978. Normal discrimination with unclassified observations. *Journal of the American Statistical Association*, 73, 821-826.
- Ott, J. & Kronmal, R. A. 1976. Some classification procedures for multivariate binary data using orthogonal functions. *Journal of the American Statistical Association*, 71, 391-399.
- Overall, J. E., Higgins, W. & Schweinitz, A. de. 1976. Comparison of differential diagnostic discrimination for abbreviated and standard MMPI. *Journal of Clinical Psychology*, 32, 237-245.
- Petersen, C. R. & Hart, D. H. 1978. Use of multiple discriminant function analysis in evaluation of a state-wide system for identification of educationally handicapped children. *Psychological Reports*, 43, 743-755.
- Pichot, P. & Perse, J. 1952. L'application des fonctions discriminantes au diagnostic individuel en psychologie. *Revue de Psychologie Appliquée*, 2, 19-34.
- Pickrel, E. W. 1958. Classification theory and techniques. *Educational and Psychological Measurement*, 18, 37-46.
- Porebski, O. R. 1966a. On the interrelated nature of the multivariate statistics used in discriminatory analysis. *The British Journal of Mathematical and Statistical Psychology*, 19, 197-214.
- Porebski, O. R. 1966b. Discriminatory and canonical analysis of technical college data. *The British Journal of Mathematical and Statistical Psychology*, 19, 215-236.
- Power, R. P., MacRae, K. D. & Muntz, H. J. 1974. Separation of normals, neurotics and simulating malingerers on the MPI by means of discriminant function analysis. *British Journal of Social and Clinical Psychology*, 13, 65-72.
- Power, R. P., Muntz, H. J. & MacRae, K. D. 1975. Man or machine as diagnostic tool: a comparison between clinical psychologists and discriminant function analysis. *British Journal of Social and Clinical Psychology*, 14, 413-422.
- Press, S. J. & Wilson, S. 1978. Choosing between logistic regression and discriminant analysis. *Journal of the American Statistical Association*, 73, 699-705.
- Quesenberry, C. P. & Gessaman, M. P. 1968. Nonparametric discrimination using tolerante regions. *Annals of Mathematical Statistics*, 39, 664-673.
- Raiffa, H. 1961. Statistical decision theory approach to item selection for dichotomous test and criterion variables. In: H. Solomon (Hrsg.). *Studies in item analysis and prediction*. 187-220. Stanford: University Press.

- Randles, R. H., Broffitt, J. D., Ramberg, J. S. & Hogg, R. V. 1978a. Discriminant analysis based on ranks. *Journal of the American Statistical Association*, 73, 379-384.
- Randles, R. H., Broffitt, J. D., Ramberg, J. S. & Hogg, R. V. 1978b. Generalized linear and quadratic discriminant functions using robust estimates. *Journal of the American Statistical Association*, 73, 564-568.
- Rao, C. R. 1948. The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society, Series B* 10, 159-203.
- Rao, C. R. 1949. On some problems arising out of discrimination with multiple characters. *Sankhyä*, 9, 343-366.
- Rao, C. R. 1951. Statistical inference applied to classificatory problems. Part III: The discriminant function approach in the classification of time series. *Sankhyä*, 11, 257-272.
- Rao, C. R. 1952. Advanced statistical methods in biometric research. Darien, Connecticut: Hafner.
- Rao, C. R. 1954a. A general theory of discrimination when the information about alternative population distributions is based on samples. *Annals of Mathematical Statistics*, 25, 651-670.
- Rao, C. R. 1954b. On the use and interpretation of distance functions in statistics. *Bulletin of the International Statistical Institute*, 34 (2), 90-97.
- Rao, C. R. 1966a. Discriminant function between composite hypotheses and related Problems. *Biometrika*, 53, 339-345.
- Rao, C. R. 1966b. Covariance adjustment and related problems in multivariate analysis. In: P. R. Krishnaiah (Hrsg.). *Multivariate analysis*. 87-103. New York: Academic Press.
- Rao, C. R. 1969. Recent advances in discriminatory analysis. *Journal of the Indian Society of Agricultural Statistics*, 21, Nr. 1, 3-15.
- Rao, C. R. 1970. Inference on discriminant function coefficients. In: R. C. Bose, I. M. Chakravarti, P. C. Mahalanobis, C. R. Rao & K. J. C. Smith (Hrsg.). *Essays in probability and statistics*. 587-602. Chapel Hill: University of North Carolina Press.
- Rao, C. R. & Slater, P. 1949. Multivariate analysis applied to differences between neurotic groups. *British Journal of Psychology, Statistical Section*, 2, 17-29.
- Rao, C. R. & Varadarajan, V. S. 1963. Discrimination of Gaussian processes. *Sankhyä, Series A* 25, 303-330.
- Rao, M. M. 1963-1964. Discriminant analysis. *Annals of the Institute of Statistical Mathematics*, 15, 11-24.
- Repges, R. 1975. Ein sequentielles nichtparametrisches Trennverfahren. *EDV in Medizin und Biologie*, 6, 9-13.
- Richards, L. E. 1972. Refinement and extension of distribution-free discriminant analysis. *Applied Statistics*, 21, 174-176.

- Rioux, P. 1975. Quadratic discriminant analysis. EDV in Medizin und Biologie, 6, 42-44.
- Rogers, G. & Linden, J. D. 1973. Use of multiple discriminant function analysis in the evaluation of three multivariate grouping techniques. Educational and Psychological Measurement, 33, 787-802.
- Rogers, W. H. & Wagner, T. J. 1978. A finite sample distribution-free performance bound for local discrimination rules. Annals of Statistics, 6, 506-514.
- Rulon, P. J. 1951. Distinctions between discriminant and regression analyses and a geometric interpretation of the discriminant function. Harvard Educational Review, 21, 80-90.
- Rulon, P. J., Tiedeman, D. V., Tatsuoka, M. M. & Langmuir, C. R. 1967. Multivariate statistics for personnel classification. New York: Wiley.
- Samuel, E. 1963. Note on a sequential classification problem. Annals of Mathematical Statistics, 34, 1095-1097.
- Saxena, A. K. 1967. A note on classification. Annals of Mathematical Statistics, 38, 1592-1593.
- Schaafsma, W. & Vark, G. N. van. 1977. Classification and discrimination problems with applications, Part I. Statistica Neerlandica, 31, 25-45.
- Schaafsma, W. & Vark, G. N. van. 1979. Classification and discrimination problems with applications, Part IIa. Statistica Neerlandica, 33, 91-126.
- Schmid, J. 1950. A comparison of two procedures for calculating discriminant function coefficients. Psychometrika, 15, 431-434.
- Schmidt, L. R. & Cattell, R. B. 1972. Differentialdiagnosen mit Hilfe objektiver Persönlichkeitstests: Diskriminanzanalytische Untersuchungen zur Depression, Manie, Schizophrenie und Neurose. Diagnostica, 18, 61-86.
- Sheslow, D. V. & Erickson, M. T. 1975. Analysis of activity preference in depressed and nondepressed college students. Journal of Counseling Psychology, 22, 329-332.
- Siotani, M. & Wang, R. H. 1977. Asymptotic expansions for error rates and comparison of the W-procedure and the Z-procedure in discriminant analysis. In: P. R. Krishnaiah (Hrsg.). Multivariate analysis - IV. 523-545. Amsterdam: North-Holland Publishing Company.
- Sitgreaves, R. 1952. On the distribution of two random matrices used in classification procedures. Annals of Mathematical Statistics, 23, 263-270.
- Sitgreaves, R. 1961. Some results on the distribution of the W-classification statistic. In: H. Solomon (Hrsg.). Studies in item analysis and prediction. 241-251. Stanford: University Press.
- Sitgreaves, R. 1973. Some operating characteristics of linear discriminant functions. In: T. Cacoullos (Hrsg.). Discriminant analysis and applications. 365-374. New York: Academic Press.
- Skarabis, H. 1970. Mathematische Grundlagen und praktische Aspekte der Diskrimination und Klassifikation. Würzburg: Physica.

- Skene, A. M. 1978. Discrimination using latent structure models. In: L. C. A. Corsten & J. Hermans (Hrsg.). COMPSTAT 1978. 199-204. Wien: Physica.
- Smith, H. F. 1936-1937. A discriminant function for plant selections. *Annals of Eugenics*, 7, 240-250.
- Smith, S. E. & Yau, S. S. 1972. Linear sequential pattern classification. *IEEE Transactions on Information Theory* IT-18, 672-678.
- Smith, W. B. & Zeis, C. D. 1973. On classification for incomplete multinormal data. *Communications in Statistics*, 2, 85-93.
- Sorum, M. 1972. Three probabilities of misclassification. *Technometrics*, 14, 309-316.
- Srivastava, J. N. & Zaatar, M. K. 1972. On the maximum likelihood classification rule for incomplete multivariate samples and its admissibility. *Journal of Multivariate Analysis*, 2, 115-126.
- Srivastava, M. S. 1967a. Comparing distances between multivariate populations - the problem of minimum distance. *Annals of Mathematical Statistics*, 38, 550-556.
- Srivastava, M. S. 1967b. Classification into multivariate normal populations when the population means are linearly restricted. *Annals of the Institute of Statistical Mathematics*, 19, 473-478.
- Stahmann, R. F. & Wallen, N. E. 1966. Multiple discriminant prediction of major field of study. *Educational and Psychological Measurement*, 26, 439-444.
- Streit, F. 1977. Identification rules based on partial information on the Parameters. In: J. R. Barra, F. Brodean, G. Romier & B. van Cutsem (Hrsg.). *Recent developments in statistics*. 797-806. Amsterdam: North-Holland Publishing Company.
- Subrahmaniam, K. & Chinganda, E. F. 1978. Robustness of the linear discriminant function to nonnormality: Edgeworth series distribution. *Journal of Statistical Planning and Inference*, 2, 79-91.
- Swircinsky, D. P. & Warnock, J. K. 1977. Comparison of the neuropsychological key and discriminant analysis approaches in predicting cerebral damage and localization. *Journal of Consulting and Clinical Psychology*, 45, 808-814.
- Tallis, G. M. 1970. Some extensions of discriminant function analysis. *Metrika*, 15, 86-91.
- Tatsuoka, M. M. & Tiedeman, D. V. 1954. Discriminant analysis. *Review of Educational Research*, 24, 402-420.
- Teichroew, D. & Sitgreaves, R. 1961. Computation of an empirical sampling distribution for the W-classification statistic. In: H. Solomon (Hrsg.). *Studies in item analysis and prediction*. 252-275. Stanford: University Press.
- Tiedeman, D. V. 1951. The utility of the discriminant function in psychological and guidance investigations. *Harvard Educational Review*, 21, 71-80.
- Toussaint, G. T. 1974. Bibliography on estimation of misclassification. *IEEE Transactions on Information Theory* IT-20, 472-479.
- Trampisch, H. J. 1975. Trennprobleme bei unvollständiger Information - Eine Übersicht. *EDV in Medizin und Biologie*, 6, 2-8.

- Trampisch, H. J. 1976. A discriminant analysis for qualitative data with interactions. *Computer Programs in Biomedicine*, 6, 50-60.
- Trampisch, H. J. 1977. Grundbegriffe der Diskriminanzanalyse. *Metamed*, 1, 365-373.
- Trampisch, H. J. 1978. Classical discriminant analysis and Lancaster models for qualitative data. In: L. C. A. Corsten & J. Hermans (Hrsg.). *COMPSTAT 1978*. 205-211. Wien: Physica.
- Travers, R. M. W. 1939. The use of a discriminant function in the treatment of psychological group differences. *Psychometrika*, 4, 25-32.
- Uematsu, T. 1964. On a multidimensional linear discriminant function. *Annals of the Institute of Statistical Mathematics*, 16, 431-437.
- Urbakh, V. Y. 1971. Linear discriminant analysis: Loss of discriminating power when a variate is omitted. *Biometrics*, 27, 531-534.
- Victor, N. 1971. A nonlinear discriminant analysis. *Computer Programs in Biomedicine*, 2, 36-50.
- Victor, N. 1976a. Probleme der Auswahl geeigneter Zuordnungsregeln bei unvollständiger Information, insbesondere für kategoriale Daten. *Biometrics*, 32, 571-585.
- Victor, N. 1976b. Non-parametric allocation rules. In: F. T. de Dombal & F. Gremy (Hrsg.). *Decision making and medical care*. 515-529. Amsterdam: North-Holland Publishing Company.
- Victor, N., Trampisch, H. J. & Zentgraf, R. 1974. Diagnostic rules for qualitative variables with interactions. *Methods of Information in Medicine*, 13, 184-186.
- Wahl, P. W. & Kronmal, R. A. 1977. Discriminant functions when covariances are unequal and sample sizes are moderate. *Biometrics*, 33, 479-484.
- Wald, A. 1944. On a statistical problem arising in the classification of an individual into one of two groups. *Annals of Mathematical Statistics*, 25, 145-162.
- Wald, A. 1947. *Sequential analysis*. New York: Wiley.
- Walter, G. F. & Porges, S. W. 1976. Heart rate and respiratory responses as a function of task difficulty: The use of discriminant analysis in the selection of psychologically sensitive physiological responses. *Psychophysiology*, 13, 563-571.
- Wani, J. K. & Kabe, D. G. 1972. Note on a multidimensional linear discriminant function. *Annals of the Institute of Statistical Mathematics*, 24, 409-412.
- Ware, J. E. & Williams, R. G. 1977. Discriminant analysis of student ratings as a means for identifying lecturers who differ in enthusiasm or information-giving. *Educational and Psychological Measurement*, 37, 627-639.
- Webster, H. 1952. Rao's multiple discriminant technique applied to three TAT variables. *Journal of Abnormal and Social Psychology*, 47, 641-648.
- Weiner, J. M. & Dunn, O. J. 1966. Elimination of variates in linear discrimination Problems. *Biometrics*, 22, 268-275.
- Welch, B. L. 1939. Note on discriminant functions. *Biometrika*, 31, 218-220.

- Wheeler, L. 1963. Predictions of brain damage from an aphasia screening test; an application of discriminant functions and a comparison with a non-linear method of analysis. *Perceptual and Motor Skills*, 17, 63-80.
- Whitney, A. W. 1971. A direct method of nonparametric measurement selection. *IEEE Transactions on Computers* C-20, 1100-1103.
- Winter, B. B. 1974. Nonparametric density estimation and statistical discrimination. *Psychological Bulletin*, 81, 371-379.
- Woinsky, M. N. & Kurz, L. 1969. Sequential nonparametric two-way classification with a prescribed maximum error probability. *Annals of Mathematical Statistics*, 40, 445-455.
- Wolde-Tsadik, G. & Yu, M. C. 1979. Concordance in variable-subset discriminant analysis. *Biometrics*, 35, 641-644.
- Yarbrough, C. 1971. Sequential discrimination with likelihood ratios. *Annals of Mathematical Statistics*, 42, 1339-1347.
- Zedeck, S. 1971. Identification of moderator variables by discriminant analysis in a multipredictable group. *Journal of Applied Psychology*, 55, 364-371.
- Zielezny, M. & Dunn, O. J. 1975. Cost evaluation of a two-stage classification procedure. *Biometrics*, 31, 37-47.

## 6. Kapitel

# Latente Strukturanalyse

***Joachim Krauth***

### 1. Einführung

Ausgangspunkt sei ein Fragebogen mit  $n$  binären Ja-Nein-Items, der von einer Stichprobe von Probanden beantwortet wird. Jedem Probanden entspricht dann ein Antwortmuster aus Einsen und Nullen, das den Mustern der Ja- und Nein-Antworten entspricht. Für jedes Muster wird die relative Häufigkeit bestimmt, mit der es aufgetreten ist.

Das Ziel einer ***latenten Klassenanalyse*** besteht darin, die Verteilung der Antwortmuster durch das Vorhandensein von zwei oder mehr ***latenten*** Klassen zu erklären. Man geht also davon aus, daß die Probandenstichprobe nicht homogen ist, sondern bezüglich einer möglicherweise nicht direkt meßbaren ***latenten Variablen*** variiert. Unterschiedliche latente Klassen entsprechen dann unterschiedlichen ***Werten*** der latenten Variablen.

Die latenten Klassen werden beschrieben durch eine Menge von latenten ***Parametern***. Derartige Parameter geben für jedes Item und jede Klasse die Wahrscheinlichkeit für die Antwort „Ja“ an. Man macht die Voraussetzung, daß man die Wahrscheinlichkeiten für die ***manifesten Variablen*** kennt, d.h. die Verteilung für die verschiedenen Antwortmuster. Mit Hilfe der sogenannten ***Berechnungsgleichungen*** (accounting equations) bestimmt man aus den manifesten Wahrscheinlichkeiten die latenten Wahrscheinlichkeiten. Um die Gleichungen aufstellen zu können, benötigt man das sogenannte ***Prinzip*** der ***lokalen Unabhängigkeit***.

Um sich die Grundidee klarzumachen, ist es gut, an ein einfaches Beispiel aus der beschreibenden Statistik anzuknüpfen, durch welches man Studienanfängern die Möglichkeit von Fehlinterpretationen von empirischen Korrelations- und Regressionskoeffizienten erläutern kann. Ausgangspunkt sei eine Zufallsstichprobe aus den Bewohnern einer Stadt. Für jeden dieser Bewohner seien die zwei Merkmale Schuhgröße und Einkommen erhoben worden, also Merk-

male, zwischen denen absolut kein Kausalzusammenhang unterstellt werden kann. Tatsächlich ergibt sich jedoch eine positive Korrelation der beiden Merkmale. Die Erklärung wird sofort einsichtig, wenn man die Meßwertpaare in eine Punktwolke (Scatter-Diagramm) einträgt und die Datenpunkte für Männer und Frauen unterschiedlich kennzeichnet. Man sieht dann einerseits, daß innerhalb der Gruppe der Männer kein Zusammenhang zwischen Schuhgröße und Einkommen besteht. Das gleiche gilt für die Gruppe der Frauen. In der Graphik ergibt sich jeweils eine ungefähr kreisförmige Punktwolke für jedes Geschlecht. Andererseits haben in der Regel Männer ein höheres Einkommen und eine höhere Schuhgröße als Frauen. Dieses führt dazu, daß die kreisförmigen Punktwolken für die beiden Geschlechter bei Fallenlassen der Unterscheidung der Geschlechter zu einer ovalen Punktwolke verschmelzen und eine positive Korrelation zwischen Schuhgröße und Einkommen vortäuschen. Falls man also eine Korrelation vorliegen hat, die man auf das Wirken einer latenten Variablen zurückführen will, wie in diesem Fall auf das Geschlecht, so muß man die vorliegende Punktwolke in kreisförmige Teilbereiche zerlegen, d.h. Klassen suchen, für die die betrachteten Merkmale lokal unabhängig sind.

In der Praxis hat man dabei natürlich größere Schwierigkeiten als in unserem Beispiel, da man den Einfluß von Geschlecht, Alter, Intelligenz usw., d.h. der offensichtlichen latenten Variablen, meist leicht berücksichtigen kann. Die latente Strukturanalyse stellt sich die Aufgabe, gerade solche nicht offensichtlichen latenten Variablen aufzuspüren.

Wie Madansky (1968) ausführt, ist die Grundidee der latenten Strukturanalyse schon sehr alt. Schon bei Cournot (1838), Weinberg (1902), Benini (1928) und de Meo (1934) werden Modelle diskutiert, mit denen man auf nicht beobachtbare (latente) Variablen mit Hilfe von polytomen beobachtbaren (manifesten) Variablen rückschließen will. Bei diesen Modellen wird die multivariate Verteilung der manifesten Variablen dargestellt als eine Mischung von multivariaten Verteilungen. Dabei ist die zu bestimmende Mischungsverteilung gerade die Verteilung der latenten Variablen.

Eine Teilkasse dieser Modelle sind diejenigen, bei denen man annimmt, daß die Variablen innerhalb jeder der Komponenten der Mischverteilung als unabhängig anzusehen sind. In Lazarsfeld (1950a) wird für solche Modelle der Begriff **latente Strukturmodelle** eingeführt. Anschließend an die beiden grundlegenden Arbeiten von Lazarsfeld (1950a, 1950b) wurden von dem gleichen Autor später noch andere Einführungs- und Übersichtsartikel verfaßt (Lazarsfeld, 1954, 1955, 1959). Einen systematischen Überblick mit einer Einbettung in die Skalierungstheorie findet man bei Torgerson (1958, S. 360-395). Einen Kurzüberblick gibt Madansky (1968).

Einführungen in deutscher Sprache geben von der Lippe (1973), Fischer (1974, S. 160-179) und als Kurzüberblick Wottawa (1979, S. 51-58).

Eine umfassende Darstellung der Theorie mit einer ausführlichen Biographie bis 1967 enthält die Monographie von Lazarsfeld & Henry (1968). Der Übersichtsartikel von Fielding (1977) berücksichtigt Literatur bis 1973.

## **2. Grundbegriffe der latenten Strukturanalyse**

Die hier vorkommenden Probleme treten immer dann auf, wenn die Variablen, die einen interessieren, nicht direkt meßbar sind und man andere Variablen messen muß, um aus den Ergebnissen Rückschlüsse auf die interessierenden Variablen ziehen zu können. Beispielsweise mag ein Psychiater bei einem Klienten das Vorhandensein oder Nichtvorhandensein gewisser Symptome beobachten und daraus Rückschlüsse darauf ziehen wollen, ob der Klient schizophren ist oder nicht. Schizophrene Klienten zeigen dabei nicht unbedingt immer alle Symptome, während sich manche Symptome auch bei nicht-schizophrenen Klienten zeigen können. Mit Hilfe der latenten Strukturanalyse versucht man nun die latenten Klassen der Schizophrenen und Nichtschizophrenen zu identifizieren.

Die einfachste Form der Analyse geht von einer Menge von dichotomen Ja-Nein-Items aus, wobei man diese Voraussetzung auch zu polytomous Items abschwächen kann. Man postuliert die Existenz eines **latenten Raumes**, in dem sich die Mitglieder einer Population befinden, wobei die Position eines Mitgliedes völlig die Wahrscheinlichkeit für eine Ja-Antwort für irgendein Item bestimmt. Der latente Raum ist somit gerade der Raum, auf dem die uns interessierende latente Variable definiert ist. Je nach der spezifischen Struktur des latenten Raumes und der Variabilität der **Itemwahrscheinlichkeiten** innerhalb dieses Raumes sprechen wir von einem speziellen **latenten Strukturmodell**. Der latente Raum kann definiert werden als diejenige Klassifikation, die für die statistischen Interaktionen zwischen den manifesten Variablen verantwortlich ist. Durch diese Klassifikation wird eine gegebene Population in homogene Teilpopulationen ‚entmischt‘ (Lazarsfeld, 1959).

Falls der latente Raum nur aus einer endlichen Anzahl von Punkten besteht, so spricht man von **diskreten Klassenmodellen**. Im obigen Beispiel würde bei n Symptomen, die jeweils nur vorhanden oder nicht vorhanden sein können, der latente Raum aus  $2^n$  möglichen Punkten bestehen. Falls wir in dem Anfangsbeispiel nur zwischen hohen (1) und niedrigen (0) Einkommen bzw. Schuhgrößen unterscheiden würden, so würde der latente Raum aus den  $2^2 = 4$  Punkten (0,0), (0,1), (1,0) und (1,1) bestehen. Ein mögliches Modell würde dann gegeben durch

$$p_{00} = p_{11} = 0.5, p_{01} = p_{10} = 0.$$

In diesem Modell würde demnach die Kombination hoher Schuhgröße mit hohem Einkommen und niedriger Schuhgröße mit niedrigem Einkommen mit

Wahrscheinlichkeit 0.5 auftreten, während die beiden Kombinationen, bei denen niedrige Werte der einen Variablen mit hohen Werten der anderen Variablen zusammentreffen, nur die Wahrscheinlichkeit 0 haben.

Falls man einen Test aus  $n$  dichotomen Items verwendet, um eine Fähigkeit zu messen, so kann man davon ausgehen, daß die latente Fähigkeit stetig verteilt ist, so daß der latente Raum z.B. durch die reelle Achse beschrieben werden kann. Die wahren Fähigkeitswerte der Probanden können dann z.B. als normalverteilt angesehen werden. Für jedes Item und jeden Punkt der Geraden, d.h. für jeden Fähigkeitswert, gibt es dann eine Wahrscheinlichkeit für eine Ja-Antwort. Diese Wahrscheinlichkeiten für jeden Punkt der Geraden kann man durch eine Funktion, die **Itemcharakteristifunktion** oder **Spurfunktion** (trace-line) beschreiben. In diesem Fall spricht man von einem **stetigen Modell**. Die Unterscheidung zwischen stetigen und diskreten Modellen bezieht sich also auf die Struktur des latenten Raumes und nicht auf die Struktur der Items, die in beiden Fällen diskrete Ausprägungen haben.

Um weitere Grundbegriffe einführen zu können, nehmen wir an, daß wir in unserem Schizophreniebeispiel insgesamt  $n = 10$  Symptome beobachten, jeweils mit den Ausprägungen vorhanden (1) und nichtvorhanden (0). Weiterhin nehmen wir an, daß es zwei Klassen von Probanden gibt, nämlich Schizophrene (1) und Nichtschizophrene (2). Das **Axiom der lokalen Unabhängigkeit** besagt:

Innerhalb jeder der beiden latenten Klassen (indiziert durch 1 und 2) sind die Reaktionen auf verschiedene Items unabhängig. Die klasseninterne Wahrscheinlichkeit jedes Reaktionsmusters für jede Teilmenge von Items ist das Produkt der entsprechenden eindimensionalen Randwahrscheinlichkeiten.

Dieses führt zu einer gewissen Anzahl von Bedingungsgleichungen, z.B.

$$p_{13}^{(1)} = p_{1\cdot}^{(1)} p_{3\cdot}^{(1)}, \quad p_{13}^{(1)} = p_{1\cdot}^{(1)} p_{3\cdot}^{(1)} = (1 - p_{1\cdot}^{(1)}) p_{3\cdot}^{(1)}, \quad p_{125}^{(2)} = (1 - p_{1\cdot}^{(2)}) p_{2\cdot}^{(2)} p_{5\cdot}^{(2)} \text{ usw.}$$

Dabei bedeutet die erste Gleichung, daß sich innerhalb der Klasse 1 der Schizophrenen aufgrund der Unabhängigkeit des Auftretens von Symptom 1 und Symptom 3 die gemeinsame Wahrscheinlichkeit als Produkt der Randwahrscheinlichkeiten ergibt. Die zweite Gleichung bedeutet, daß die Wahrscheinlichkeit dafür, daß Symptom 1 nicht auftritt, während gleichzeitig Symptom 3 auftritt, sich innerhalb der Klasse 1 der Schizophrenen als Produkt der entsprechenden Randwahrscheinlichkeiten ergibt. Schließlich wird in der dritten Gleichung die Wahrscheinlichkeit für das gleichzeitige Nichtauftreten von Symptom 1 und das Auftreten von Symptom 2 und Symptom 5 innerhalb der Klasse 2 der Nichtschizophrenen als Produkt der entsprechenden Randwahrscheinlichkeiten dargestellt.

Über das Axiom der bedingten lokalen Unabhängigkeit und seine Bedeutung auch für andere Skalierungsansätze informieren Lord & Novick (1968, S. 538-540). Es zeigt sich, daß es schwer ist, auf diese Annahme zu verzichten. In der Faktorenanalyse kann man dieses Postulat zwar auf die lineare lokale Unabhängigkeit abschwächen, darf dafür aber auch nur Interaktionen 1. Ordnung zulassen. Harper (1972) diskutiert, wie man die lokale Unabhängigkeit wenigstens soweit abschwächen kann, daß eine paarweise Abhängigkeit zwischen den Items zugelassen ist.

Bei  $n$  dichotomen Items wird ein Zweiklassenmodell durch  $2n + 2$  Parameter charakterisiert. Für die erste Klasse sind dies die  $n$  latenten Wahrscheinlichkeiten bzw. latenten Randwahrscheinlichkeiten  $p_1^{(1)}, \dots, p_n^{(1)}$ , für die zweite Klasse die  $n$  latenten Wahrscheinlichkeiten  $p_1^{(2)}, \dots, p_n^{(2)}$ . Zusätzlich benötigt man noch die Wahrscheinlichkeit  $v^{(1)}$  bzw.  $v^{(2)}$ , mit der ein Proband zu der Klasse 1 bzw. der Klasse 2 gehört. Die Summe der Klassenwahrscheinlichkeiten ergibt immer 1:

$$1 = v^{(1)} + v^{(2)}.$$

Dieses ist die erste Modellgleichung. Die weiteren Berechnungsgleichungen ergeben sich mit Hilfe der Formel von der totalen Wahrscheinlichkeit, denn es gilt z.B. für die Wahrscheinlichkeit  $p_1$  für das Auftreten des Symptoms 1 in der Population

$$p_1 = p_1^{(1)}v^{(1)} + p_1^{(2)}v^{(2)}.$$

Dieses liegt daran, daß z.B. der unbekannte latente Parameter  $p_2^{(1)}$  die bedingte Wahrscheinlichkeit dafür ist, daß ein Proband, der zu Klasse 1 gehört, das Symptom 2 aufweist. Entsprechend ergibt sich z.B. die Wahrscheinlichkeit dafür, daß ein Proband in der Population das Merkmal 1 aufweist und das Merkmal 3 nicht aufweist, zu

$$p_{13} = p_{13}^{(1)}v^{(1)} + p_{13}^{(2)}v^{(2)} = p_1^{(1)}(1 - p_3^{(1)})v^{(1)} + p_1^{(2)}(1 - p_3^{(2)})v^{(2)}.$$

Beobachtbar sind nur die relativen Häufigkeiten der verschiedenen Auftretenskombinationen der Symptome. Diese verwendet man als Schätzungen für die entsprechenden Wahrscheinlichkeiten. So ist die relative Häufigkeit der Personen, bei denen Symptom 1 und Symptom 5 nicht auftreten, jedoch das Symptom 3, eine Schätzung für die entsprechende nichtbedingte Wahrscheinlichkeit  $p_{135}$ . Indem man diese Schätzungen anstelle der unbekannten Wahrscheinlichkeiten in die Berechnungsgleichungen einsetzt, bekommt man ein Gleichungssystem für die unbekannten latenten Parameter  $v^{(1)}, v^{(2)}, p_1^{(1)}, \dots, p_n^{(1)}, p_1^{(2)}, \dots, p_n^{(2)}$ . Durch Auflösen dieses Gleichungssystems erhält man Schätzungen für diese latenten Parameter.

Wie man leicht-einsieht, ergeben sich bei  $n$  dichotomen Items und 2 latenten Klassen insgesamt  $2^n$  Berechnungsgleichungen (Lazarsfeld & Henry, 1968, S.

27). Für  $n=2$  Items gibt es  $2^2=4$  Gleichungen für  $2n+2=2 \cdot 2 + 2 = 6$  Parameter. In diesem Fall kann man die Parameter und damit auch das Modell nicht **identifizieren**. Da in den Gleichungen die gesamte Information über das Modell enthalten ist, besteht keine Möglichkeit, eine eindeutige Lösung zu finden, es sei denn, wir erlegen den latenten Parametern noch weitere Restriktionen auf.

Für  $n=3$  Items gibt es  $2^3=8$  Gleichungen für  $2 \cdot 3 + 2 = 8$  Parameter. Hier besteht zumindest die Möglichkeit, eine eindeutige Lösung zu erhalten. Bei der Lösung dieses nichtlinearen Gleichungssystems geht man zunächst von zwei Annahmen aus. Man nimmt an, daß das spezielle latente Strukturmodell mit zwei Klassen tatsächlich vorliegt und daß man die **manifesten** nichtbedingten Wahrscheinlichkeiten auf den linken Seiten der Gleichungen tatsächlich kennt. Dann reduziert sich das Problem auf eine rein algebraische Fragestellung. Die Schätzung der latenten Parameter, falls man nur Schätzungen der manifesten Wahrscheinlichkeiten hat, und die Anpassungsgüte des latenten Strukturmodells an die beobachteten Daten sind eine danach zu behandelnde Frage.

Die algebraische Frage der Lösung der Gleichungssysteme findet man behandelt in Lazarsfeld (1961) und in Lazarsfeld & Henry (1968). Für die manifesten Wahrscheinlichkeiten werden dazu sogenannte **Kreuzprodukte** der Items i und j

$$[ij] = p_{ij} - p_i p_j$$

und sogenannte **geschichtete Kreuzprodukte**

$$[ij; k] = p_{ijk} p_k - p_{ik} p_{jk}$$

eingeführt.

Mit Hilfe dieser Ausdrücke werden die Berechnungsgleichungen umgeschrieben und dadurch formal leichter lösbar. Die Lösungen brauchen jedoch keineswegs Werte zwischen 0 und 1 anzunehmen. Mit Hilfe der Kreuzprodukte lassen sich eine Reihe von notwendigen Bedingungen formulieren, die für ein vorgegebenes latentes Strukturmodell erfüllt sein müssen. Ein weiterer Schritt ist die Einführung sogenannter **symmetrischer Parameter**. Ein symmetrischer Parameter 3. Ordnung ist definiert durch

$$[ijk] = p_{ijk} - p_i[jk] - p_i[ik] - p_k[ij] - p_i p_j p_k.$$

Falls man mit  $X_i$  die Zufallsvariable bezeichnet, die gleich 1 ist, falls eine Ja-Antwort bei Item i vorliegt und gleich 0 sonst, so gilt für die Kovarianz

$$[ij] = E[(X_i - p_i)(X_j - p_j)] = \text{Cov}[X_i, X_j]$$

und für das zentrale Moment 3. Ordnung

$$[ijk] = E[(X_i - p_i)(X_j - p_j)(X_k - p_k)].$$

Entsprechendes gilt für symmetrische Parameter höherer Ordnung. Derartige Beziehungen werden sowohl bei Strukturuntersuchungen als auch bei der Herleitung von Schätzverfahren verwendet.

Mit Hilfe der latenten Parameter kann man für jedes mögliche Reaktionsmuster die Wahrscheinlichkeit dafür berechnen, daß ein Proband, der dieses Muster gezeigt hat, zu einer bestimmten Klasse gehört. Mit Hilfe dieser **Rekrutierungswahrscheinlichkeiten** (recruitment probabilities) kann man Probanden in der Weise klassifizieren, daß man sie der für sie wahrscheinlichsten Klasse zuordnet. Für  $n=3$  ergäbe sich z.B. für das Muster (Ja, Nein, Ja) die Wahrscheinlichkeit

$$p_{1\bar{2}3} = p_{1\bar{2}3}^{(1)} v^{(1)} + p_{1\bar{2}3}^{(2)} v^{(2)} = p_1^{(1)} (1-p_2^{(1)}) p_3^{(1)} v^{(1)} + \\ + p_1^{(2)} (1-p_2^{(2)}) p_3^{(2)} v^{(2)}.$$

Der erste Summand ist die Wahrscheinlichkeit dafür, daß ein Proband aus der Klasse 1 das Muster aufweist, der zweite Summand die entsprechende Wahrscheinlichkeit für einen Probanden der Klasse 2. Man ordnet den Probanden der Klasse mit der größeren Wahrscheinlichkeit zu. Addiert man diese größeren Wahrscheinlichkeiten über alle Muster, so erhält man die totale Wahrscheinlichkeit dafür, einen Probanden richtig zu klassifizieren. Zieht man diesen Wert von 1 ab, so ergibt sich die Wahrscheinlichkeit einer Fehlklassifikation.

Weist ein Proband das obige Muster auf, so ist die Wahrscheinlichkeit dafür, daß er dann zu der Klasse 1 gehört, gegeben durch

$$P(\text{Klasse 1} \mid \text{Muster (Ja, Nein, Ja)}) = v^{(1)} p_{1\bar{2}3}^{(1)} / p_{1\bar{2}3},$$

Für mehr als  $n=3$  Items erhalten wir mehr Berechnungsgleichungen als unbekannte latente Parameter. Z.B. ergeben sich für  $n=4$  Items  $2^4=16$  Gleichungen für  $2 \cdot 4 + 2 = 10$  Parameter. Damit handelt es sich nicht mehr um ein **Interpolationsproblem**, bei dem ein Gleichungssystem aufzulösen ist, sondern um ein **Approximationsproblem**, bei dem  $2n+2$  Parameterschätzwerte gesucht sind, die die  $2^n$  Gleichungen mit den  $2^n - 1$  beobachteten relativen Häufigkeiten möglichst gut erfüllen. Jedoch wurde in der ersten Zeit bei der Schätzung der latenten Parameter immer versucht, auch bei mehr als  $n=3$  Items die Interpolationsmethode zu verwenden, indem man aus der Menge der Gleichungen immer nur so viel Gleichungen auswählte, wie unbekannte latente Parameter zu schätzen waren (algebraische Schätzverfahren).

### **3. Allgemeines Vorgehen bei der latenten Strukturanalyse**

In Lazarsfeld (1959) wird ein Schema aus 9 Schritten diskutiert, die bei einer latenten Strukturanalyse durchzuführen sind.

Der erste Schritt betrifft die **Auswahl** und **Spezifizierung** des Modells. Dieses betrifft Annahmen über die Spurfunktionen und über die Verteilung der Population über dem latenten Raum.

Als zweiter Schritt werden die **Berechnungsgleichungen** für das spezielle Modell aufgestellt. Diese verknüpfen die Verteilung der manifesten Variablen mit der der latenten Variablen.

Der dritte Schritt betrifft die sogenannten **Reduzierbarkeitsbedingungen**. Für mehr als  $n=3$  Items erhält man mehr Gleichungen als unbekannte latente Parameter. Dieses führt dazu, daß die manifesten Wahrscheinlichkeiten gewissen zusätzlichen Restriktionen unterliegen, wenn ein bestimmtes Strukturmodell vorliegt. Diese Restriktionen haben die Form von Gleichungen, die besonders einfach mit Hilfe der Kreuzprodukte und geschichteten Kreuzprodukte formuliert werden können. Man spricht von Reduzierbarkeitsbedingungen, weil das System der Berechnungsgleichungen nur bei Erfülltsein dieser Bedingungen gelöst werden kann. Sind die Bedingungen nicht erfüllt, so ist das Modell falsch.

Im vierten Schritt muß die **Identifizierbarkeit** überprüft werden, d.h. es muß geklärt werden, ob die vorliegenden manifesten Parameter ausreichen, um die Werte der latenten Parameter festzulegen. Wie sich im dritten Schritt gezeigt hat, reicht es dazu nicht aus, zu überprüfen, ob mehr Gleichungen als Unbekannte vorliegen. Aus den Reduzierbarkeitsbedingungen geht hervor, daß viele der manifesten Parameter aus anderen manifesten Parametern berechnet werden können. Die Berechnungsgleichungen sind also nur zum Teil unabhängig.

Im fünften Schritt geht **es** um die **Identifikation** des Modells. Hierbei ist nach der tatsächlichen Auflösung der Berechnungsgleichungen nach den latenten Parametern gefragt. Man stellt bei der Identifikation die latenten Parameter dar durch Funktionen der manifesten Parameter. Die Darstellung wird wieder vereinfacht durch die Verwendung der Kreuzprodukte.

Der sechste Schritt betrifft die Anpassungsprozedur oder das **Schätzverfahren**. Die bisherigen Schritte bezogen sich alle auf algebraische Fragestellungen innerhalb eines bestimmten Modells. Bei der Anpassung konkreter Daten hat man es zusätzlich mit Stichprobenschwankungen zu tun. Während im Modell ein bestimmtes Kreuzprodukt immer dasselbe ist, in welchen Berechnungsgleichungen es auch auftreten mag, können sich empirische Kreuzprodukte, in die anstelle der Randwahrscheinlichkeiten die entsprechenden relativen Häufigkeiten eingesetzt werden, durchaus voneinander unterscheiden. Dieses führte zu dem Vorschlag, gewisse Mittelwertbildungen über die empirischen Kreuzprodukte vorzunehmen. Andere Schätzverfahren werden noch besprochen.

Im siebten Schritt ist die **Anpassungsgüte** zu überprüfen, d.h. ein Maß dafür anzugeben, wie gut die Daten zu dem angepaßten Modell passen. Da die Daten kaum jeweils die Reduzierbarkeitsbedingungen exakt erfüllen, werden die aufgrund des Modells vorhergesagten manifesten Wahrscheinlichkeiten auch nicht mit den beobachteten relativen Häufigkeiten exakt übereinstimmen. Für ein dem Problem angemessenes Modell sollten aber die Abweichungen klein und zufällig verteilt sein.

Im achten Schritt wird das **Rekrutierungsmuster** bestimmt. Damit soll die Frage angesprochen werden, wo sich ein Individuum mit einem bestimmten Antwortmuster im latenten Raum befindet. Prinzipiell kann sich die Person in jedem Punkt des latenten Raumes befinden. Im allgemeinen wird sie aber mit großer Wahrscheinlichkeit aus einem bestimmten Bereich des latenten Raumes stammen und mit geringer Wahrscheinlichkeit aus anderen Bereichen. In diesem Sinne entspricht jedem Antwortmuster ein Vektor von Aufenthaltswahrscheinlichkeiten, das Rekrutierungsmuster.

Im neunten und letzten Schritt geht es um die Frage der **Klassifikation** der Probanden bzw. um die Zuordnung von **Skalenwerten**. Man sucht dann nach einem Kriterium, um einen Probanden aufgrund seines Antwortmusters einem bestimmten Punkt des latenten Raumes, der besonders typisch für ihn ist, zuzuordnen. Gelegentlich begnügt man sich auch damit, die Antwortmuster gemäß einem solchen Kriterium anzugeben. Zur Schätzung solcher Skalenwerte reicht die Kenntnis der ersten Momente, z.B. des Erwartungswertes, der Rekrutierungsverteilung schon aus. Falls man einer Person den für sie wahrscheinlichsten Punkt im latenten Raum als Skalenwert zuordnen will, so benötigt man die Kenntnis der ganzen Rekrutierungsverteilung.

In diesem Zusammenhang interessiert man sich oft noch für ein anderes duales Problem, die **Skalierung der Items**. Man möchte den Items Skalenwerte zuordnen, die angeben, wieweit ein einzelnes Item zu den Skalenwerten der Individuen beiträgt bzw. wieweit jedes Item zwischen Individuen diskriminiert, die an verschiedenen Punkten des latenten Raumes lokalisiert sind. Diese Fragestellung wird in Lazarsfeld (1954) diskutiert.

## 4. Modelle der latenten Strukturanalyse

### 4.1 Allgemeines Modell

In Anlehnung an Anderson (1959) beschreibt Fielding (1977) das allgemeine latente Strukturmodell (siehe auch McDonald, 1962). Es seien  $X_1, \dots, X_n$  manifeste Variablen und entsprechend  $Y_1, \dots, Y_m$  latente Variablen. Die Variablen können sich prinzipiell auf beliebigem Skalenniveau befinden. Dabei ist ge-

wöhnlich m als viel kleiner als n angenommen, meist sogar m=l. Man nimmt eine bedingte Wahrscheinlichkeitsverteilung  $f(x_1, \dots, x_n | y_1, \dots, y_m)$  der X's bei gegebenen Y's an, wobei f eine Dichte ist für stetige X's und eine Menge von Wahrscheinlichkeiten für diskrete X's. Entsprechend sei eine Randverteilung der Y's gegeben durch  $g(y_1, \dots, y_m)$ . Damit folgt für die nichtbedingte Randverteilung von  $X_1, \dots, X_n$

$$h(x_1, \dots, x_n) = \int f(x_1, \dots, x_n | y_1, \dots, y_m) g(y_1, \dots, y_m) dy_1 \dots dy_m.$$

Nach der Formel von Bayes folgt für ein Individuum mit den Werten  $x_{o1}, \dots, x_{on}$  der manifesten Variablen für die bedingte Dichte bzw. bedingten Wahrscheinlichkeiten bezüglich der Werte  $y_1, \dots, y_m$  des Individuums im latenten Raum

$$k(y_1, \dots, y_m | x_{o1}, \dots, x_{on}) = \frac{f(x_{o1}, \dots, x_{on} | y_1, \dots, y_m) g(y_1, \dots, y_m)}{h(x_{o1}, \dots, x_{on})}.$$

Aufgrund von  $x_{o1}, \dots, x_{on}$  kann man mit Hilfe dieser Gleichung ein Individuum einem Punkt im latenten Raum zuweisen. Naheliegend ist die Maximum-Likelihood-Methode, nach der ein Individuum mit  $(x_{o1}, \dots, x_{on})$  den Punkt  $(y_{o1}, \dots, y_{om})$  zugewiesen bekommt, für den die Funktion k maximal wird. Falls man Fehlklassifikationen für verschiedene Punkte für unterschiedlich schwerwiegend hält, so wird man die Punkte unterschiedlich gewichten, d.h. Entscheidungsregeln mit geeigneten Verlustfunktionen verwenden.

Falls man die obige Gleichung verwenden will, muß man davon ausgehen, daß man allenfalls die Funktion h für die manifesten Variablen kennt, während f und g unbekannt sind und auch nicht eindeutig aus h gefolgert werden können. Deshalb verwendet man das **Axiom der lokalen Unabhängigkeit**, welches besagt, daß für einen festen Punkt im latenten Raum die manifesten Variablen unabhängig sind. Für m=l latente Variable bedeutet dieses

$$f(x_1, \dots, x_n | y) = f_1(x_1 | y) \dots f_n(x_n | y)$$

bzw.

$$h(x_1, \dots, x_n) = \int f_1(x_1 | y) \dots f_n(x_n | y) g(y) dy.$$

Inhaltlich bedeutet das Axiom, daß Individuen mit gleichen latenten Ausprägungen einander ähnlich sind und daß die gesamte relevante Information in den latenten Variablen liegt, so daß nach deren Festlegung jedes Verhalten nur noch zufällig ist. Man geht also davon aus, daß die sich zwischen den manifesten Variablen zeigenden Abhängigkeiten allein auf Abhängigkeiten in den latenten Variablen zurückzuführen sind. Das Axiom wird genauer diskutiert bei Lord & Novick (1968, S. 538-540). Eine Abschwächung des Axioms zumindest für das latente Klassenmodell in dem Sinne, daß die Items paarweise abhängig sein dürfen, sieht Harper (1972) vor.

Die unter den Werten der manifesten Variablen bedingten Dichten bzw. Wahrscheinlichkeiten  $f_i(x_i \mid y)$  heißen **Spurfunktionen**. Diese versucht man zusammen mit der latenten Verteilung  $g$  zu ermitteln, um die latente Struktur zu erhalten. Dann weiß man, wie die manifesten Variablen von den latenten Variablen abhängen und kann Individuen skalieren und die Parameter der Funktionen schätzen.

Darstellungen des allgemeinen latenten Strukturmodells findet man außer in den oben erwähnten Arbeiten auch in Torgerson (1958, S. 361-367) und Green (1952).

## 4.2 Existenzproblem

Bei der Konstruktion eines Modells geht man von einer manifesten Wahrscheinlichkeitsfunktion  $h(x_1, \dots, x_n)$  aus, die noch von gewissen unbekannten aber schätzbarer Parametern abhängt. Beispielsweise kann  $h$  die Dichte einer  $n$ -dimensionalen Normalverteilung sein. Ebenso wählt man die Spurfunktionen  $f_i$  und die latente Verteilung  $g$  als Wahrscheinlichkeitsfunktionen vorgegebener Form mit noch freien Parametern. Auch die Anzahl  $m$  der latenten Variablen muß spezifiziert sein. Es ist dann durchaus nicht gesagt, daß es Parameter gibt, so daß die obige Beziehung zwischen  $h$ ,  $f_1, \dots, f_n$  und  $g$  erfüllbar ist. z.B. existiert kein latentes Strukturmodell, falls  $h$  einer Normalverteilung und die  $f_i$  Cauchyverteilungen entsprechen (Anderson, 1959). Unter Umständen kann die Frage der Existenz eines latenten Strukturmodells durch die Frage der Gültigkeit gewisser Reduzierbarkeitsbedingungen beantwortet werden. Man vergleiche dazu Lazarsfeld (1959, S. 512-515).

## 4.3 Identifikationsproblem

Falls man  $h$  kennt, so reicht die durch das Axiom der lokalen Unabhängigkeit gegebene Beziehung zwischen  $h$ ,  $f_1, \dots, f_n$  und  $g$  nicht aus, um  $f_1, \dots, f_n$  und  $g$  eindeutig zu identifizieren. Man benötigt Zusatzannahmen über die Form der Spurfunktionen bzw. der latenten Verteilung, um die Struktur zu identifizieren. Je nach den speziellen Annahmen ergeben sich dann die einzelnen Modelle.

Selbst innerhalb dieser Untermodelle ist es möglich, daß verschiedene Strukturen, die verschiedenen Parameterkombinationen entsprechen, mit einer Verteilung  $f$  verträglich sind. Viele Arbeiten haben sich damit beschäftigt, Bedingungen herauszufinden, unter denen sich bei einem speziellen Modell eine Struktur, gegeben durch  $f_1, \dots, f_n$  und  $g$ , eindeutig identifizieren lässt. Dazu gehören z.B. McHugh (1956, 1958), Madansky (1960) und Goodman (1974a).

Insbesondere das allgemeine Identifikationsproblem wurde untersucht von Koopmans & Reiersol (1950) und Koopmans (1949, 1951).

Häufig ist es nicht möglich, für ein gegebenes Untermodell die Struktur zu identifizieren. Für praktische Zwecke ausreichend ist dann oft die eindeutige Identifikation gewisser Strukturaspekte, z.B. allein die Identifikation gewisser Erwartungswerte anstelle einer ganzen Verteilung (Fielding, 1977). Auch können nichtidentifizierbare Modelle durch Einschränkungen der möglichen Parameter oft identifizierbar gemacht werden (Goodman, 1974a).

#### 4.4 Strukturproblem

Falls man unter Kenntnis von  $h$  das Existenzproblem und das Identifikationsproblem gelöst hat, so sind als nächstes die zu  $f_1, \dots, f_n$  und  $g$  gehörigen Parameter zu bestimmen (Anderson, 1959). Dieses Problem ist für jedes Untermodell getrennt zu untersuchen. In der Sprechweise von Lazarsfeld (1959) ist dieses das Identifikationsproblem.

#### 4.5 Latentes Klassenmodell

Eines der einfachsten Modelle ist das latente Klassenmodell von Lazarsfeld (1950a). Hier nimmt  $Y$  eine endliche Menge von Werten, sogenannte *latente Klassen*, an. Die latente Verteilung  $g$  ist eine Multinomialverteilung, während die Spurfunktion  $f_i(x_i | y)$  die Wahrscheinlichkeit für ein Individuum aus der Klasse  $y$  ist, für das Item  $i$  die Reaktion  $x_i$  zu zeigen. Gesucht ist in diesem Falle also ein Modell mit  $q$  latenten Klassen, den Werten der latenten Variablen  $Y$ , die so geartet sind, daß innerhalb der Klassen die  $n$  manifesten Variablen (Items) unabhängig sind. Wenn solche Klassen konstruiert sind, z.B. die Klassen der Schizophrenen und Nichtschizophrenen, so sind alle Interaktionen zwischen den manifesten Variablen aufgeklärt. Auf dieses Modell (latent class model) war im Anfang schon eingegangen worden, und es ist auch in der Literatur am häufigsten betrachtet worden.

Besonderes Interesse hat das Identifikationsproblem, d.h. die Frage der eindeutigen Lösbarkeit der Berechnungsgleichungen gefunden. Bei  $n$  manifesten Variablen und  $q$  latenten Klassen ergeben sich  $2^n$  Gleichungen für  $q(n+1)$  latente Parameter. Für  $2^n > q(n+1)$  sind verschiedene Lösungsvorschläge gemacht worden. Dabei versuchte man meist durch geeignete Auswahl der Gleichungen das Ziel einer eindeutigen Parameterbestimmung zu erreichen.

Ein erster Lösungsansatz für das Strukturproblem scheint der von Green (1951) zu sein. Nach Umschreibung der Berechnungsgleichungen in Matrizenform wird analog zur Faktorenanalyse eine Lösung durch Faktorisierung von zwei Matrizen angestrebt.

Im einzelnen betrachtet Green (1951) die  $(n+l) \times (n+l)$ -Matrix  $P_0$ , deren erstes Element eine 1 ist und in deren 1. Zeile und 1. Spalte ansonsten die eindimensionalen Randwahrscheinlichkeiten stehen. Hierbei ist  $n$  die Anzahl der Items. In den übrigen Zeilen bzw. Spalten befinden sich die zweidimensionalen Randwahrscheinlichkeiten. Weiterhin wird die  $(qxq)$ -Diagonalmatrix  $V$  definiert, die in der Diagonale die Wahrscheinlichkeiten  $v^{(1)}, \dots, v^{(q)}$  enthält, die angeben, mit welcher Wahrscheinlichkeit ein zufällig ausgewählter Proband zu einer der  $q$  Klassen gehört. Schließlich sei  $L$  eine  $(n+l) \times q$ -Matrix mit den Elementen  $p_i^{(j)}$  die die bedingten Wahrscheinlichkeiten dafür bezeichnen, daß eine Person aus der latenten Klasse  $j$  das Item  $i$  positiv beantwortet.

Die Matrix  $A$ , definiert durch

$$A = LV^{1/2},$$

enthält dann die unbekannten latenten Parameter bzw. man kann diese bei Kenntnis von  $A$  leicht bestimmen. Aufgrund der Berechnungsgleichungen gilt

$$P_0 = AA',$$

wobei  $A'$  die Transponierte von  $A$  ist. Mit Hilfe eines faktorenanalytischen Ansatzes läßt sich  $P_0$  darstellen als

$$P_0 = BB',$$

wobei

$$BE = A$$

gilt mit einer orthogonalen Matrix  $E$ . Um  $E$  zu bestimmen, faktorisiert man in ähnlicher Weise die Matrix  $P_1$ , die die dreidimensionalen Randwahrscheinlichkeiten in symmetrischer Form enthält:

$$P_1 = CC'.$$

Aus der Matrix

$$TT' = (B'B)^{-1}B'C$$

ergibt sich dann bis auf die Anordnung der Items eindeutig die gesuchte Matrix  $E$  mit Hilfe einer Hauptkomponentenanalyse. Eine auf T. W. Anderson zurückgehende Vereinfachung (vgl. Green, 1951) bestimmt  $TT'$  aus

$$TT' = (B'B)^{-1}B'P_1B(B'B)^{-1}.$$

Damit erspart man sich die Faktorisierung von  $P_1$ .

Eine Schwierigkeit bereiten die Randwahrscheinlichkeiten mit wiederholt auftretenden gleichen Indizes in den Matrizen  $P_0$  und  $P_1$ . Dieses sind artifizielle

Größen ohne realen Bezug zu den manifesten Wahrscheinlichkeiten, was zu gewissen Schätzproblemen führt.

Eine andere Matrixmethode für dieses Problem wurde zuerst von Lazarsfeld & Dudman (1951) sowie unabhängig davon durch Koopmans (1951) vorgeschlagen. Diese Methode, die auf die Bestimmung der Nullstellen einer Determinantengleichung hinausläuft, wurde von Anderson (1954) weiterentwickelt. Durch Gibson (1955) wurde das Verfahren so modifiziert, daß alle Items simultan berücksichtigt wurden. Darauf aufbauend konnte Madansky (1960) das Verfahren verallgemeinern, insbesondere in Hinblick auf notwendige und hinreichende Bedingungen für die Identifizierbarkeit. Gibson (1962a) diskutierte die Möglichkeit, zusätzliche Items nachträglich in eine latente Klassenanalyse mit einzubeziehen. Die hier beschriebene Matrixmethode wird auch als **Basismethode** bezeichnet. Außer in den angegebenen Originalarbeiten wird sie in Lazarsfeld & Henry (1968, S. 46-48) und in Fielding (1977) dargestellt.

Die wesentliche Bedingung für die Identifizierbarkeit des latenten Klassenmodells besteht darin, daß die Anzahl der Gleichungen mindestens so groß ist wie die Anzahl der Unbekannten, d.h.  $2^n \geq q(n+1)$ . Für eine hinreichend große Anzahl von manifesten Variablen sollte eine Lösung immer dadurch möglich sein, daß man nur Gleichungen auswählt, die sich auf Randverteilungen niedriger Dimensionen beziehen, z.B. höchstens der zweiten Dimension. In diesem Fall würden also nur die Berechnungsgleichungen für die  $p_i$  und die  $p_{ij}$  berücksichtigt. Alternativ könnte man weniger manifeste Variablen und dafür Randverteilungen höherer Dimensionen betrachten (Gibson, 1955).

Bei der Basismethode muß man aus der Menge der  $n$  manifesten Variablen drei disjunkte Teilmengen auswählen, von denen zwei je  $q-1$  und eine genau eine Variable enthält. Dabei sei  $q$  die Anzahl der latenten Klassen. Die eine ausgezeichnete Variable bezeichnet man auch als **Schichtungsvariable** (stratifier). Man stellt die ausgewählten Berechnungsgleichungen auf die folgende Weise in Matrixform dar. Auf der linken Seite steht eine Matrix  $B_{hv}$ , die sogenannte **Basismatrix**. Diese ist von der Form, daß in der ersten Spalte die eindimensionalen Randwahrscheinlichkeiten für die positive Beantwortung eines Items der ersten Teilmenge stehen und in der ersten Zeile die entsprechenden Größen für die zweite Teilmenge. Im Schnittpunkt, d.h. im ersten Element der Matrix steht eine Eins. Die restlichen Elemente der Matrix geben die zugehörigen zweidimensionalen Randwahrscheinlichkeiten wieder. Diese  $(q \times q)$ -Matrix läßt sich aufgrund der zugehörigen Berechnungsgleichungen darstellen in der Form

$$B_{hv} = L_h V L_v'$$

Hier ist  $L_h$  eine Matrix, deren Zeilen jeweils die latenten Parameter für die Items aus der ersten Teilmenge enthalten, und  $L_v$  die entsprechende Matrix für

die Items aus der zweiten Teilmenge. Zusätzlich enthalten  $L_h$  und  $L_v$  ein weiteres artificielles Item an erster Stelle, für das alle latenten Wahrscheinlichkeiten Eins sind, d.h. die erste Zeile dieser Matrizen enthält nur Einsen. Dieses führt dazu, daß  $L_h$  und auch  $L'_v$ , die Transponierte von  $L_v$ ,  $(q \times q)$ -Matrizen sind. Mit  $V$  wird wieder eine  $(q \times q)$ -Diagonalmatrix bezeichnet, die in der Diagonale die Wahrscheinlichkeiten  $v^{(0)}, \dots, v^{(q)}$  enthält. Die Aufgabe besteht darin, die Matrizen  $L_h$ ,  $V$  und  $L'_v$  zu bestimmen, da diese die unbekannten latenten Parameter enthalten. Dazu stellt man eine weitere Matrizen-Gleichung auf, gegeben durch

$$B_{hv;k} = L_h V D_k L'_v.$$

Hier ist  $B_{hv;k}$  die **geschichtete Basismatrix**. Diese ergibt sich aus der Basismatrix  $B_{hv}$  durch **Schichtung** bezüglich des **Schichtungsitems** (stratifier item), das hier durch  $k$  indiziert wird. Die Schichtung erfolgt so, daß in  $B_{hv}$  jede eindimensionale Randwahrscheinlichkeit durch eine entsprechende zweidimensionale Randwahrscheinlichkeit ersetzt wird, indem man als zweite Variable die Schichtungsvariable hinzunimmt. Entsprechend werden die zweidimensionalen Randwahrscheinlichkeiten durch dreidimensionale ersetzt. Das erste Element von  $B_{hv}$ , eine Eins, wird durch  $p_k$  ersetzt, d.h. durch die Wahrscheinlichkeit für eine positive Beantwortung des Schichtungsitems. Die Matrix  $D_k$  ist eine  $(q \times q)$ -Diagonalmatrix, die in der Diagonale die latenten Wahrscheinlichkeiten des Schichtungsitems und sonst Nullen enthält. Es ergibt sich, daß die Lösungen  $t$  der Determinantengleichung

$$|L_h D_k L_h^{-1} - tI| = 0,$$

wobei  $I$  die Einheitsmatrix ist, gerade die latenten Wahrscheinlichkeiten für das Schichtungsitem sind. Da  $B_{hv}$  und  $B_{hv;k}$  als bekannt vorausgesetzt werden und

$$L_h D_k L_h^{-1} = B_{hv;k} B_{hv}^{-1}$$

gilt, erhält man auf diese Weise die latenten Parameter für das Schichtungsitem. Aufgrund der Beziehungen

$$(B_{hv;k} B_{hv}^{-1} - I) p_k^{(i)} L_h^{(i)} = 0, i=1, \dots, q,$$

wobei  $L_h^{(i)}$  der  $i$ -te Spaltenvektor von  $L_h$  ist, erhält man durch Auflösung dieser homogenen linearen Gleichungssysteme die Matrix  $L_h$ . Die Lösungen dieser Gleichungssysteme sind eigentlich nur bis auf konstante Faktoren eindeutig bestimmt, jedoch wird Eindeutigkeit dadurch erzwungen, daß das erste Element von  $L_h^{(i)}$  immer 1 sein muß. Ferner gilt

$$V L'_v = L_h^{-1} B_{hv}.$$

Da V eine Diagonalmatrix ist und die erste Spalte von  $L'_v$  aus Einsen besteht, enthält die erste Spalte von  $VL'_v$  die Klassenwahrscheinlichkeiten  $v^{(1)}, \dots, v^{(q)}$ . Falls man auf diese Weise V bestimmt hat, ergibt sich schließlich  $L'_v$  aus

$$L'_v = V^{-1} L_h^{-1} B_{hv}.$$

Liegen mehr als  $(q-1) + (q-1) + 1 = 2q-1$  Items vor und bezeichnet r die Menge dieser zusätzlichen Items, so kann man die nicht notwendig quadratische Matrix  $P_{hr}$  aufstellen, die die manifesten zweidimensionalen Randwahrscheinlichkeiten  $p_{ij}$  enthält, wobei i zu h und j zu r gehört. Die fehlenden latenten Parameter ergeben sich dann aus

$$L'_r = V^{-1} L_h^{-1} P_{hr}.$$

Es ergeben sich vier Bedingungen, die erfüllt sein müssen, damit ein latentes Klassenmodell mit q Klassen bei n Items mit Hilfe der Basismethode eindeutig gelöst werden kann (Lazarsfeld & Henry, 1968, S. 56):

- (1) Es muß  $n \geq 2q-1$  gelten.
- (2) Alle Klassen müssen nicht leer sein.
- (3) Es gibt mindestens ein Item, das unterschiedliche latente Wahrscheinlichkeiten hat. Dieses wird als Schichtungselement verwendet.
- (4) Es gibt zwei disjunkte Teilmengen der Items aus jeweils  $q-1$  Elementen, die das Schichtungselement nicht enthalten. Diese Mengen sind so geartet, daß die Matrizen  $L_h$  und  $L_v$  Inverse haben.

Praktische Kritikpunkte an der Basismethode bestehen darin, daß sie wesentlich von der Auswahl der  $2q-1$  Items und deren Zerlegung in drei disjunkte Teilmengen abhängt. Dieses führt zu einer Abhängigkeit von der Anordnung der Items. Weiterhin ist nicht garantiert, daß die Lösungen reelle Zahlen zwischen 0 und 1 sind.

Eine Erweiterung des obigen Ansatzes beruht auf der Verwendung der sogenannten **anstiegenden Matrizen** (ascending matrices) (Lazarsfeld & Henry, 1968, S. 60-62), die nicht von den Items sondern von den Antwortmustern her definiert werden. Dabei werden Randverteilungen höherer Ordnung mit einbezogen und weniger manifeste Variablen benötigt (Madansky, 1960).

Von Gibson (1955) wurde eine zusammengesetzte Schichtungsvariable vorgeschlagen. Dieses bedeutet, daß in der Diagonalen der Matrix  $D_k$  nicht mehr die latenten Wahrscheinlichkeiten eines einzelnen Items sondern gewichtete oder ungewichtete Summen der latenten Wahrscheinlichkeiten von mehreren Schichtungsvariablen stehen. Damit sollen die Lösungen der Determinanten-Gleichung besser getrennt werden können.

Einen alternativen Ansatz bei Vorliegen von zwei manifesten Variablen, der auf der sogenannten **singular value decomposition** einer Matrix beruht, geben

Good (1969) und Gilula (1979) an. Speziell leitet Gilula (1979) notwendige und hinreichende Bedingungen für die Existenz von dichotomen latenten Variablen, eine Technik zur Identifikation der Verteilungen dieser Variablen sowie die bedingten Verteilungen der manifesten Variablen bei gegebenen latenten Variablen her.

Die Probleme, die auftreten, falls man die ursprünglichen Items informationsmäßig äquivalenten Transformationen unterwirft und dann latente Klassenmodelle anpaßt, diskutieren Batchelder & Narens (1977). Eine Spezifikation des latenten Klassenmodells zum Zwecke der Itemanalyse bei Leistungstests betrachtet Wilcox (1979a, 1979b).

## 4.6 Latentes Polynommodell

Man geht von einer stetigen latenten Variablen  $Y$  aus, die über einem gewissen Intervall definiert ist, und nimmt an, daß ein Individuum mit dem Wert  $y$  der latenten Variablen bei Item  $i$  die Reaktion  $x_i = 1$  mit der Wahrscheinlichkeit

$$f_i(1 | y) = a_{0i} + a_{1i}y + a_{2i}y^2 + \dots + a_{ri}y^r$$

zeigt. Der Grad  $r$  des Polynoms, durch das die Spurfunktion beschrieben wird, soll bekannt sein in der gleichen Weise wie die Anzahl der Klassen im latenten Klassenmodell. Die Wahl dieses Grades wird zweckmäßigerweise so getroffen, daß die Anpassungsgüte des Modells an die Daten hinreichend gut ist. Die latente Verteilung  $g$  wird oft als Beta-Dichte angenommen. Dieses Polynommodell (latent polynomial model) wird ausführlich in Lazarsfeld & Henry (1968, S. 197-217) behandelt.

Auf den ersten Blick erscheint es sinnlos, Spurfunktionen als Polynome zu wählen, weil die bedingten Wahrscheinlichkeiten  $f_i(1 | y)$  Zahlen zwischen 0 und 1 sein müssen. Dieses ist sicher nicht erfüllt, falls man Polynome betrachtet, die über der ganzen reellen Achse definiert sind. Deshalb nimmt man an, daß die latente Variable  $y$  nur über einem gewissen Intervall  $\alpha < y < \beta$  absolut stetig verteilt ist, d.h. dort eine Wahrscheinlichkeitsdichte besitzt, und daß die  $f_i(1 | y)$  Zahlen zwischen 0 und 1 sind für  $y$  aus diesem Intervall. Dieses führt zu gewissen Restriktionen für die Polynome, d.h. zu Restriktionen für die Polynomkoeffizienten. Die letzteren werden gelegentlich, um die Darstellungen zu erleichtern, als von Null verschieden angenommen. Dieses ist aber keine notwendige Voraussetzung.

Die Vorteile des Polynommodells bestehen darin, daß man durch Polynome viele stetige Spurfunktionen approximieren kann und daß die manifesten Wahrscheinlichkeiten nur von den Momenten der latenten Verteilung abhängen.

Eine Auflösung der Berechnungsgleichungen ergibt sich wieder über eine Matrixdarstellung. Die Basismatrix  $B_{hv}$  der manifesten Wahrscheinlichkeiten wird wieder als Produkt dreier Matrizen dargestellt:

$$B_{hv} = A_h' M A_v.$$

Hier sind  $A_h$  und  $A_v$  Matrizen der jeweiligen Polynomkoeffizienten und  $M$  ist eine Momentenmatrix. Die Auflösung wird dadurch erschwert, daß die Matrix  $D_k$  für die Schichtungsvariable keine Diagonalmatrix mehr ist sondern eine ***Semi-Diagonalmatrix***, in der neben der Hauptdiagonalen auch Nachbar diagonalen besetzt sind.

Speziell sind lineare und quadratische Spurfunktionen untersucht worden, weil sowohl die Lösungsmethoden als auch die Interpretationsmöglichkeiten hier am einfachsten sind. Darstellungen hierfür findet man bei Lazarsfeld & Henry (1968, S. 206-211) und speziell für lineare Spurfunktionen bei Torger son (1958, S. 367-374).

## 4.7 Lokalisiertes Klassenmodell

Man geht wieder von polynomia len Spurfunktionen aus, nimmt man aber an, daß die latente Verteilung  $g$  eine diskrete Verteilung mit einer endlichen Anzahl von Stützstellen ist. Es liegt also ein latentes Kontinuum vor, auf dem die Klassen an gewissen Punkten lokalisiert sind (located class model). Falls die Anzahl der Stützstellen höchstens gleich  $r-1$  ist ( $r = \text{Grad der Polynome}$ ), so ergibt sich wieder das latente Klassenmodell. Im anderen Fall bestehen Einschränkungen für die Werte der Spurfunktionen. Dieses Modell wird in Lazarsfeld & Henry (1968, S. 148-156) behandelt.

Die Annahme diskreter Klassen und einer stetigen Skala mag widersprüchlich erscheinen. Jedoch kann es einerseits sein, daß diskrete angeordnete Klassen existieren, über deren Abstand man Genaueres wissen möchte. Die andere Interpretation wäre, daß die Klassenstruktur eine Approximation an eine stetige Skala ist.

Ebenso wie im Polynommodell sind auch im lokalisier ten Klassenmodell insbesondere die linearen (Lazarsfeld & Henry 1968, S. 149-153) und quadratischen (Lazarsfeld & Henry 1968, S. 153-156) Spurfunktionen untersucht worden.

## 4.8 Latentes Inhaltsmodell

In diesem Modell verwendet man anstelle der Polynome die Funktionen

$$f_i(1 | y) = a_i + b_i y d_i$$

und nimmt für die latente Dichte  $g$  eine Gleichverteilung über dem Intervall  $(0,1)$  an. Eine Darstellung des latenten Inhaltsmodells (latent content model) wird gegeben in Lazarsfeld & Henry (1968, S. 160-181). Ausführlicher ist Somers (1961).

Im latenten Inhaltsmodell wird angenommen, daß die latente Variable  $Y$  nur Werte zwischen 0 und 1 annehmen kann. Damit auch die bedingten Wahrscheinlichkeiten  $f_i(1 | y)$  zwischen 0 und 1 liegen, müssen die Koeffizienten  $a_i$ ,  $b_i$  und  $d_i$  folgenden Restriktionen unterliegen:

$$d_i \geq 0, 0 < a_i < 1, 0 < a_i + b_i < 1.$$

Außerdem ist es zweckmäßig

$$b_i > 0$$

zu fordern, damit mit Wachsendem  $y$  die Wahrscheinlichkeit für die positive Beantwortung eines Items wächst. Diese Wachstumsrate kann von Item zu Item in Abhängigkeit von  $d_i$  unterschiedlich sein.

Das latente Inhaltsmodell geht von einer stetigen latenten Variablen aus, und man kann deshalb annehmen, daß sich Datensätze, für die dieses Modell geeignet ist, wesentlich von Datensätzen unterscheiden, für die das latente Klassenmodell anzuwenden ist. Die Lösung der Berechnungsgleichungen erfordert für dieses Modell allerdings einige neue Überlegungen. Hierbei gibt es Probleme mit nicht eindeutigen Lösungen.

Wegen der auftretenden Schwierigkeiten wurden einige Spezialfälle genauer untersucht: 1. Der Fall gleicher Krümmungen, d.h. gleicher Werte  $d_i$ ; 2. der Fall, daß die Spurfunktionen aller Items in Null beginnen, d.h. alle Werte  $a_i$  sind gleich Null; 3. der Fall von symmetrischen Spurfunktionen (vgl. Lazarsfeld & Henry 1968, S. 191-194).

Als Alternative zu der stetigen Gleichverteilung der latenten Variablen wird in Lazarsfeld & Henry (1968, S. 195-196) auch die Betaverteilung betrachtet. Damit soll eine stärkere Gewichtung für einen Punkt der Skala (Modalwert der Betaverteilung) erreicht werden.

## 4.9 Latentes Distanzmodell

Im latenten Distanzmodell (latent distance model) wird ähnlich wie im lokализierten Klassenmodell versucht, eine gewisse Anordnung der Klassen zu berücksichtigen. Es handelt sich um eine probabilistische Version der sogenannten **Guttman-Skalierung**. In Torgerson (1958, S. 374) wird darauf hingewiesen, daß auch jedes andere Modell mit monotonen Spurfunktionen und hinreichend gut diskriminierenden Items als Analogon zur Guttman-Skalierung dienen kann. In dem deterministischen Modell von Guttman wird angenommen, daß falls ein Individuum ein Item  $i$  richtig beantwortet und das nächstschwere Item  $i+1$  falsch beantwortet, es alle Items, die leichter als  $i$  sind, richtig beantwortet und alle Items, die schwerer als  $i+1$  sind, falsch beantwortet.

Die latente Variable  $Y$  sei stetig verteilt und

$$f_i(1 \mid y) = a_i \text{ für } y \leq c_i, \quad f_i(1 \mid y) = b_i \text{ für } y > c_i,$$

wobei gilt

$$c_1 < c_2 < \dots < c_n, \quad a_i < b_i.$$

Das so definierte latente Distanzmodell ist ein eingeschränktes latentes Klassenmodell. Ein Proband wird für einen sehr niedrigen Wert von  $y$  alle Items mit hoher Wahrscheinlichkeit falsch beantworten. Für Werte  $y$  mit  $c_1 < y \leq c_2$  wird er Item 1 mit hoher Wahrscheinlichkeit richtig und die anderen Items falsch beantworten, usw. Im deterministischen Grenzfall mit  $a_i = 0$ ,  $b_i = 1$  ergibt sich das Guttman-Modell.

In Lazarsfeld & Henry (1968, S. 140-142) und in Hays & Borgatta (1954) wird auch der Spezialfall mit der Restriktion  $a_i = 1 - b_i$  näher untersucht. Das latente Distanzmodell wird allgemein in Lazarsfeld & Henry (1968, S. 123-148), in Hays & Borgatta (1954), in Torgerson (1958, S. 374-385) und Proctor (1970) diskutiert.

## 4.10 Testtheoretisches Modell

Es sei  $Y$  eine stetige latente Variable und  $f_i(1 \mid y)$  die Dichte einer  $N(a_i, b_i^2)$ -Verteilung. Man nimmt an, daß  $Y$  gemäß  $N(0,1)$  verteilt ist. Eine andere Formulierung des Modells geht von einer Darstellung (Anderson, 1959)

$$Z = cY + a + U$$

aus, wobei  $Z$ ,  $c$ ,  $a$  und  $U$  Vektoren der Dimension  $n$  sind,  $U_i$  normalverteilt mit der Varianz  $b_i^2$  und  $Y$  gemäß  $N(0,1)$  verteilt ist. Dieses ist ein Faktorenanalysemodell mit einem gemeinsamen Faktor. Dieses Modell wird in der psycholo-

gischen Testtheorie verwendet, wobei ein Alternativmodell anstelle der Normalverteilung die logistische Verteilung verwendet.

Diskutiert werden die testtheoretischen Modelle, darunter auch das **Rasch-Modell**, von Lazarsfeld & Henry (1968, S. 217-225). Eine Gegenüberstellung von latenter Strukturanalyse und Testtheorie findet man bei Lazarsfeld (1959, 1960).

Während man das testtheoretische Modell in der obigen Form in die latente Strukturanalyse einbetten kann, werden in der Testtheorie gewisse Modifikationen des Modells vorgenommen, um es den dort gestellten Anforderungen besser anzupassen. So bemerken Lord & Novick (1968, S. 358-394) in ihrem Vergleich mit den latenten Strukturmodellen, daß anders als in der latenten Strukturanalyse in der Testtheorie nicht alle Antwortmuster sondern nur gewisse Statistiken dieser Muster, die die für testtheoretische Fragestellungen wichtigen Informationen enthalten, betrachtet werden. So wird z.B. anstelle des Vektors der Einzelantworten nur die Statistik ‚Anzahl der richtigen Antworten‘ verwendet, die unter gewissen Voraussetzungen suffizient ist, d.h. die gesamte relevante Information enthält.

Eine Anwendung der latenten Strukturanalyse in der Testtheorie gibt Merdith (1965) aufbauend auf McDonald (1962). In der Arbeit wird unter anderem die empirische Bestimmung von Spurfunktionen und die Schätzung der Item-reliabilitäten diskutiert.

In Torgerson (1958, S. 385-395) wird das testtheoretische Modell auf der Basis von Lord (1953) und Tucker (1952) diskutiert. Dabei ergeben sich das Normal-Ogiven-Modell von Tucker (1952) und das Intelligenzskalenmodell von Thurstone (1925) als Spezialfälle eines allgemeinen Normal-Ogiven-Modells.

#### 4.11 Latentes Profilmodell

Im latenten Profilmodell (latent profile model) geht man von  $q$  latenten Klassen, d.h.  $q$  diskreten Werten der latenten Variablen  $Y$  aus, nimmt aber an, daß die manifesten Variablen  $X_1, \dots, X_n$  stetig verteilt sind. Für die bedingten Verteilungen der  $X_i$  werden keine speziellen Annahmen gemacht, aber gefordert, daß einige der bedingten Momente, z.B.  $E[X_i | y_j]$  für  $i=1, \dots, n$ ,  $j=1, \dots, q$ , eindeutig spezifiziert sind. Für jede manifeste Variable erhält man auf diese Weise einen Vektor der bedingten Erwartungswerte bei festen Klassen

$$(E[X_i | y_1], \dots, E[X_i | y_q]),$$

der als **latentes Profil** bezeichnet wird. Damit eine Struktur identifizierbar ist, muß man aus der als bekannt vorausgesetzten gemeinsamen Verteilung der

$X_1, \dots, X_n$  das Profil, den Vektor der bedingten Varianzen und die durch  $v^{(1)}, \dots, v^{(4)}$  gegebene latente Verteilung bestimmen.

Die latente Profilanalyse ergibt für jedes Item ein Profil, d.h. gibt für jedes Item an, welchen Erwartungswert es in jeder der latenten Klassen hat. Man wendet zur Lösung der Berechnungsgleichungen die Basismethode auf die standardisierten manifesten Variablen an und erhält deshalb als Lösung nicht das latente Profil sondern das standardisierte latente Profil. Durch Multiplikation mit der Standardabweichung und Addition des Erwartungswertes für das jeweilige Item ergibt sich aber sofort das nichtstandardisierte latente Profil.

Das latente Profilmmodell wird besprochen bei Lazarsfeld & Henry (1968, S. 228-239) und Fielding (1977). In Gibson (1959) erfolgt die Verallgemeinerung der latenten Klassenanalyse zur latenten Profilanalyse, um Beziehungen zwischen quantitativen Variablen zu untersuchen. McDonald (1962) untersucht den Zusammenhang mit dem latenten Klassenmodell. In Takane (1976) werden für die bedingten Verteilungen der manifesten Variablen bei festen latenten Klassen speziell Normalverteilungen angenommen. Eine Verallgemeinerung des latenten Profilmmodells derart, daß nicht nur die manifesten sondern auch die latenten Variablen stetig verteilt sind, wird in Lazarsfeld & Henry (1968, S. 239-244) versucht.

## 4.12 Andere Modelle

Eine naheliegende Verallgemeinerung des latenten Klassenmodells mit dichotomen Items ist die Erweiterung auf polytome Items. Man vergleiche hierzu Lazarsfeld & Henry (1968, S. 226-228) und Formann (1978b). In diese Richtung gehören auch die Ansätze, die die latente Klassenanalyse in die Theorie der multivariaten Kontingenztafeln einbetten. Beispiele dafür sind die Arbeiten von Goodman (1974a, 1974b, 1978) und Haberman (1974, 1977). In Goodman (1974b) ist vor allem der Ansatz der **Pfaddiagrammanalyse** bemerkenswert, während die anderen Artikel die Theorie der **loglinearen Modelle** verwenden.

Eine andere Erweiterung der Theorie betrifft den Einsatz von latenten Klassenmodellen bei der Analyse von Verläufen, d.h. in Meßwiederholungsplänen. Hier wurde ein latentes **Markov-Ketten-Modell** in Lazarsfeld & Henry (1968, S. 252-263) betrachtet. Man nimmt dabei an, daß die Population zu jedem Zeitpunkt in latente Klassen zerfällt und daß die Antwortwahrscheinlichkeiten nur durch diese Klassen bestimmt werden. Zusätzlich sollen sich die Individuen entsprechend einer homogenen Markovkette von Klasse zu Klasse bewegen können. Bei  $q$  latenten Klassen führt dieses zu einer  $(qxq)$ -Matrix der Übergangswahrscheinlichkeiten. Ein ähnliches, wenn auch nicht identisches Modell schlug Coleman (1964) vor. Einen weiteren Ansatz zu einer

latenten Markov-Ketten-Analyse mit einer Anwendung auf das Gefangenendilemma entwickelt van der Sanden (1978).

## 5. Statistische Fragestellungen

### 5.1 Parameterschätzung

#### 5.1.1 Einführung

Der ursprüngliche Ansatz der latenten Strukturanalyse war algebraischer und nicht statistischer Natur. Man ging davon aus, daß die gemeinsame Verteilung  $h(x_1, \dots, x_n)$  der manifesten Variablen bekannt ist und daß das speziell betrachtete Modell die tatsächliche latente Struktur richtig wiedergibt. Das Problem der Bestimmung der latenten Parameter bestand somit vor allem darin, Bedingungen anzugeben und Gleichungen auszuwählen, die eine Identifikation der Parameter ermöglichen. Bei der Anwendung der latenten Strukturanalyse kann man jedoch weder davon ausgehen, daß man die Verteilung der manifesten Variablen kennt, noch daß das angenommene Modell exakt zutrifft. Die verfügbaren Daten weichen aufgrund von Meß- und Stichprobenfehlern von den zu schätzenden Parametern ab und sind als Realisationen von manifesten Zufallsvariablen anzusehen. Es besteht also das Bedürfnis nach Schätzverfahren für die latenten Parameter, die gewissen Gütekriterien genügen.

#### 5.1.2 Algebraische Verfahren

Bei den sogenannten **algebraischen Verfahren** ersetzt man die unbekannten ein- und mehrdimensionalen manifesten Randwahrscheinlichkeiten  $p_i$ ,  $p_{ik}$ ,  $p_{ijk}$  usw. durch die entsprechenden relativen Häufigkeiten und verwendet dann die für die Parameter entwickelten algebraischen Methoden, um die Berechnungsgleichungen nach den unbekannten latenten Parametern aufzulösen. Weil auf den linken Seiten der Gleichungen jetzt nicht mehr die manifesten Randwahrscheinlichkeiten stehen, sondern nur deren Schätzungen, erhält man als Lösungen ebenfalls nur Schätzungen der latenten Parameter.

Schwierigkeiten entstehen dadurch, daß verschiedene Bedingungen an die manifesten Parameter gestellt werden müssen, damit die algebraischen Techniken verwendet werden können. Es handelt sich hierbei um Bedingungen, die die Identifizierbarkeit der Parameter und die Invertierbarkeit gewisser Matrizen sicherstellen. Eine erste Voraussetzung ist also die Wahl von Items, deren zugehörige Lösungswahrscheinlichkeiten die angesprochenen Bedingungen erfüllen. Eine zweite Voraussetzung betrifft den vorgegebenen Datensatz. Auch die vorgefundenen relativen Häufigkeiten müssen die Bedingungen er-

füllen, d.h. dürfen nicht allzuweit von den zugehörigen Wahrscheinlichkeiten abweichen. Dieses setzt im Verhältnis zu der Anzahl der betrachteten Items große Stichprobenumfänge voraus. Jedoch muß man in jedem Falle mit einer gewissen Wahrscheinlichkeit damit rechnen, daß die Verfahren versagen, denn auch bei noch so großen Stichprobenumfängen besteht eine gewisse positive Wahrscheinlichkeit dafür, daß z.B. alle Versuchspersonen die Items in der gleichen Weise beantworten oder sich eine sonstige ungünstige Konstellation ergibt, die keinen Rückschluß auf die latenten Parameter zuläßt.

Eine weitere Schwierigkeit ist die Möglichkeit unzulässiger Lösungen. Die Verfahren lassen nämlich durchaus zu, daß man negative oder Eins übersteigende Schätzungen für die latenten Wahrscheinlichkeiten erhält. Selbst nichtreelle Zahlen sind u.U. als Lösungen möglich.

Schließlich kann man nicht unbedingt damit rechnen, eindeutige Lösungen zu erhalten. Je nachdem, welche Gleichungen oder welche Schichtungsvariablen man auswählt, kann man unterschiedliche Lösungen erhalten. Man benötigt deshalb zusätzliche Selektionskriterien. Solche Kriterien werden diskutiert in Anderson (1959) und darauf basierend in Fielding (1977). So erfordert die Basislösung die Inversion von Matrizen, deren Determinanten demnach von Null verschieden sein müssen. Es scheint deshalb naheliegend zu sein, bei der Basismethode die Itemmenge so in die drei Teilmengen von manifesten Variablen zu unterteilen, daß die Determinanten groß werden. Dieses Vorgehen scheinen Simulationsstudien zu bestätigen. Ein Vorschlag ist, diejenigen  $M$  derartigen Zerlegungen auszuwählen, die die größten Werte der Determinanten liefern. Für jede der  $M$  Zerlegungen ist dann das Gleichungssystem zu lösen, wobei nichtzulässige Lösungen ausgeschieden werden. Schließlichmittelt man über alle zulässigen Lösungen.

Ein Alternativvorschlag besagt, daß man für jede Lösung ein Maß für die Anpassungsgüte des Modells bestimmt und dann die Lösung mit der besten Anpassung aussucht. Dazu vergleicht man etwa die beobachteten Häufigkeiten mit den aufgrund der Parameterschätzungen durch das Modell vorhergesagten Wahrscheinlichkeiten mit Hilfe eines Chiadratkriteriums. Die Lösung mit dem kleinsten Chiadratwert wird dann ausgewählt. Einen solchen Ansatz findet man bei Lazarsfeld & Henry (1968, S. 98).

Die sich bei den algebraischen Verfahren ergebenden Punktschätzungen sind **konsistent**. Dieses bedeutet, daß man durch genügend große Wahl des Stichprobenumfanges immer erreichen kann, daß die Wahrscheinlichkeit dafür, daß sich Schätzung und zu schätzender Parameter um mehr als eine vorgegebene Differenz (absolut genommen) unterscheiden, kleiner als eine vorgegebene Wahrscheinlichkeit ist. Auch bei einer konsistenten Schätzung kann es also mit einer gewissen Wahrscheinlichkeit vorkommen, daß sich Parameter und Schätzung auch bei großem Stichprobenumfang stark unterscheiden. Jedoch ist

diese Wahrscheinlichkeit bei konsistenten Punktschätzungen klein. In diesem Sinne sind relative Häufigkeiten konsistente Punktschätzungen für Wahrscheinlichkeiten. Sofern die Schätzungen für die latenten Parameter stetige Funktionen der relativen Häufigkeiten sind, erhält man auch konsistente Schätzungen für die latenten Parameter. Außerdem sind diese Schätzungen asymptotisch normalverteilte Zufallsvariablen. Für die latente Klassenanalyse findet man eine Herleitung von T. W. Anderson in Lazarsfeld & Henry (1968, **S. 273-287**).

Das bisher angesprochene algebraische Verfahren verwendet nur die Schätzungen der manifesten Randwahrscheinlichkeiten  $p_i$ ,  $p_{ij}$  und  $p_{ijk}$ , d.h. nur die ersten drei Momente, und wurde ursprünglich von Lazarsfeld & Dudman (1951) vorgeschlagen und von Anderson (1954, 1959) erweitert. Es beruht im Wesentlichen auf der Bestimmung der Nullstellen einer Determinantengleichung. Mehr Information nutzt ein Ansatz von Gibson (1955) aus. Eine Verallgemeinerung durch Mitverwendung von höheren Momenten stammt von Madansky (1958, 1960). In Gibson (1962a) wird angegeben, wie die latenten Parameter von Items zu schätzen sind, die nachträglich hinzugenommen wurden und bei denen man dieselbe latente Struktur unterstellen kann.

### 5.1.3 Faktorisierungsmethoden

Von Gibson (1951, 1956) wurde ein Ansatz vorgeschlagen, der nur die Berechnungsgleichungen für die Randwahrscheinlichkeiten 1. und 2. Ordnung berücksichtigt und das Gleichungssystem durch Faktorisierung einer Matrix analog zu den Verfahren der Faktorenanalyse zu lösen versucht. Eine kurze Darstellung gibt Anderson (1959). Ein Vorteil dieses Verfahrens ist darin zu sehen, daß es keine Auswahl von Berechnungsgleichungen voraussetzt wie im Falle der algebraischen Methoden. Dafür kommt eine Nichteindeutigkeit der Lösungen dadurch zustande, daß das Ergebnis einer solchen Faktorisierung nur bis auf orthogonale Rotationen eindeutig bestimmt ist. Hier liegt es nahe, die Berechnungsgleichungen höherer als 2. Ordnung mit heranzuziehen, um eine eindeutige Lösung zu erhalten. Ein Algorithmus hierfür wurde von Green (1970) für die latente Klassenanalyse vorgeschlagen, der dann von Mardberg (1971) und Isaacson (1972) für die latente Profilanalyse verwendet wurde. Der ursprüngliche Ansatz von Green (1951), der auf der Berücksichtigung der 3. Momente beruht, wird auch von Anderson (1959) diskutiert.

### 5.1.4 Maximum-Likelihood-Schätzungen

Sowohl die algebraischen Methoden als auch die Faktorisierungsmethoden bei der Schätzung der latenten Parameter verwenden ein in der Statistik unübliches

zweistufiges Vorgehen. Zunächst werden die manifesten Verteilungen als bekannt angesehen und eine algebraische Lösung des Gleichungssystems für die latenten Parameter angestrebt. Im zweiten Schritt werden dann in die Lösungsgleichungen für die latenten Parameter die beobachteten relativen Häufigkeiten anstelle der manifesten Wahrscheinlichkeiten eingesetzt. Dieses Verfahren führt zwar zu konsistenten und asymptotisch normalverteilten Schätzungen der latenten Parameter, wird aber in der Statistik deshalb kaum verwendet, weil es andere Verfahren gibt, die ebenfalls konsistente und asymptotisch normalverteilte Schätzungen liefern, die aber eine größere Effizienz haben. Damit ist gemeint, daß diese Verfahren für große Stichprobenumfänge Schätzungen liefern, die eine kleinere Varianz als die Schätzungen des obigen Verfahrens haben, d.h. Schätzungen die weniger um den zu schätzenden Parameter streuen. Falls man Schätzungen hat, die konsistent und asymptotisch normalverteilt sind, sowie asymptotisch die kleinstmögliche Varianz haben, so spricht man von BAN-Schätzern (best asymptotic normal). Derartige Schätzer liefert unter bestimmten Voraussetzungen die **Maximum-Likelihood-Methode**. Diese geht von einem bekannten Verteilungstyp für die manifesten Variablen aus, z.B. einer n-dimensionalen Multinomialverteilung (McHugh, 1956, 1958), und versucht Schätzungen der latenten Parameter so zu bestimmen, daß die Wahrscheinlichkeit dafür, daß gerade die vorgefundene relative Häufigkeiten auftreten, maximal wird. Man kann auch sagen, daß zwischen dem Vektor der relativen Häufigkeiten und dem Vektor der zugehörigen Wahrscheinlichkeiten ein gewisser Abstand besteht und man die Schätzungen so wählt, daß falls man sie anstelle der Wahrscheinlichkeiten in das Abstandsmaß einsetzt, dieser Abstand minimal wird. Bei den Maximum-Likelihood-Schätzungen ist der Abstand gegeben durch die „Unwahrscheinlichkeit“ mit der ein Vektor von relativen Häufigkeiten bei einem vorgegebenen Vektor von Wahrscheinlichkeiten auftritt. Andere Schätzverfahren, die ebenfalls zu BAN-Schätzern führen, verwenden als Abstandsmaß eine Chiquadratstatistik (Minimum-Chiquadrat-Methode, modifizierte Minimum-Chiquadrat-Methode) oder eine Informationsstatistik.

Bei der Maximum-Likelihood-Methode verfährt man so, daß man die **Likelihoodfunktion**, d.h. die Wahrscheinlichkeitsfunktion, in die man die vorgefundenen Meßwerte eingesetzt hat, nach den unbekannten Parametern partiell differenziert und die Ableitungen gleich Null setzt. Diese sogenannten **Maximum-Likelihood-Gleichungen** sind dann nach den Schätzungen der gesuchten Parameter aufzulösen. Da eine explizite Lösung meist nicht möglich ist, verwendet man Iterationsverfahren. Die Ausgangswerte liefert dabei meist eines der oben angesprochenen Verfahren (Henry, 1966).

Ein Vorteil der Maximum-Likelihood-Methode ist sicher, daß die Willkürlichkeit der Lösung in Abhängigkeit von den ausgewählten Gleichungen nicht mehr auftritt. Jedoch setzt das Verfahren die Kenntnis der latenten Verteilung

und der Spurfunktionen voraus, was außer im Falle der latenten Klassenanalyse zu Schwierigkeiten führen kann.

Für die latente Klassenanalyse findet man die Herleitung der Schätzungen bei McHugh (1956) mit einer Korrektur in McHugh (1958). In Formann (1978a, 1978b) wird durch eine vorausgehende Transformation der latenten Parameter erreicht, daß diese zwischen Null und Eins liegen müssen. Die transformierten Parameter werden dann mit der Maximum-Likelihood-Methode geschätzt. Man erhält mit diesem Verfahren zwar immer zulässige Lösungen, jedoch im allgemeinen keine Maximum-Likelihood-Schätzungen für die ursprünglichen Parameter (vgl. Brunk, 1958). In Goodman (1979b) wird im Gegensatz zu einer Behauptung von Formann (1978a) bewiesen, daß das Iterationsverfahren von Goodman (1974a, 1974b) zu Maximum-Likelihood-Schätzungen führt, die immer innerhalb des zulässigen Intervalls liegen.

Erweiterungen auf latente Strukturmodelle mit polytomen Items findet man bei Goodman (1974a, 1974b) und Haberman (1974, 1977). Für die latente Profilanalyse unter der Annahme, daß die bedingten Verteilungen der manifesten Variablen bei gegebenen latenten Klassen Normalverteilungen sind, findet man bei Takane (1976) Maximum-Likelihood-Schätzer hergeleitet.

### **5.1.5 Aufteilungsmethoden**

Von Madansky (1958, 1959) wurde ein Verfahren vorgeschlagen, das direkt die Hauptvoraussetzung der latenten Strukturanalyse, nämlich die lokale Unabhängigkeit, ausnutzt und Schätzungen der latenten Parameter danach bewertet, wie gut dieses Axiom erfüllt ist. Im latenten Klassenmodell ordnet man jedes der N Individuen genau einer der q latenten Klassen zu, d.h. teilt die Stichprobe in q Teilstichproben auf. Für jede Teilstichprobe muß das Axiom der lokalen Unabhängigkeit erfüllt sein. Man kann dann für jede Teilstichprobe die Teststatistik eines Chiquadrattests auf Unabhängigkeit berechnen und diese Größen über alle q Klassen aufsummieren (Anderson, 1959). Diese Summe bildet ein Maß der bedingten Unabhängigkeit über alle Klassen. Eine Schätzprozedur besteht dann darin, unter allen möglichen Aufteilungen diejenige auszuwählen, die den kleinsten Wert der Summenchiquadratstatistik liefert. Dieses Minimum ist nicht notwendig asymptotisch chiquadratverteilt. Die latenten Parameter werden dann durch die üblichen in der Chiquadrat-Statistik verwendeten Schätzwerte geschätzt.

Ein numerisch einfacheres Maß wurde von Madansky (1958, 1959) anstelle der Chiquadratsumme vorgeschlagen (vgl. Anderson, 1959; Fielding, 1977). Die Minimierung dieses Maßes führt zu konsistenten Schätzungen für die latenten Parameter, und es lassen sich auch asymptotische Varianzen dieser Schätzer angeben.

Da die Anzahl der möglichen Aufteilungen mit wachsenden N und q sehr schnell anwächst, scheint das Verfahren nur für relativ kleine Stichprobenumfänge und wenige latente Klassen praktikabel zu sein. In Cassady, Miller & Dingman (1968) wird eine Version des Verfahrens betrachtet, die weniger Rechenzeit benötigt. Ein anderer Ansatz könnte darin bestehen, eine rechentechnisch noch zu bewältigende Menge von Zufallsaufteilungen zu verwenden und eines der Kriterien in einer solchen Menge zu minimieren.

### **5.1.6 Andere Schätzmethoden**

Andere naheliegende Schätzverfahren wurden schon im Zusammenhang mit der Maximum-Likelihood-Methode diskutiert. Dieses waren die Chiquareat-Minimum-Methode, die als Abstandsmaß die Chiquaretanpassungsstatistik verwendet (Lazarsfeld & Henry, 1968, S. 99-105; van der Waerden, 1971), ferner die modifizierte Chiquareat-Minimum-Methode, bei der man den Nenner der Anpassungsstatistik vereinfacht (van der Waerden, 1971) und die von Madansky (1958) verwendet wurde sowie die auf der Informationsstatistik beruhenden Schätzer (Kuliback, 1959). Diese Verfahren führen ebenso wie die Maximum-Likelihood-Methode zu BAN-Schätzern. Eine Entscheidung für eines der Verfahren wäre demnach vor allem unter dem Gesichtspunkt des Rechenaufwandes zu treffen. Leider führen alle diese Verfahren zu komplizierten Schätzern.

Ein iteratives Verfahren von J. E. Moyal beschreibt Anderson (1959). Man wählt eine beliebige Zuordnung der Individuen zu den latenten Klassen. Diese führt direkt zu Schätzungen für die latenten Parameter durch relative Häufigkeiten. Aufgrund dieser Schätzungen ordnet man jedes Individuum der Klasse zu, für die die geschätzte bedingte Wahrscheinlichkeit am größten ist. Aufgrund dieser Zuordnung erhält man neue Schätzungen, usw. Dieses Verfahren ist zwar rechentechnisch einfach, führt aber gemäß Anderson (1959) zu nicht-konsistenten Schätzungen.

In Takane (1972) wird das latente Klassenmodell als Spezialfall der Faktorenanalyse interpretiert und die Rekrutierungswahrscheinlichkeiten analog zu der Schätzung der Faktorenwerte geschätzt. Diese Schätzungen werden mit den direkten Schätzungen der Rekrutierungswahrscheinlichkeiten verglichen.

### **5.1.7 Programme und Algorithmen**

Angesichts der Probleme, die das Schätzen der latenten Parameter mit sich bringt, ist es nicht verwunderlich, daß seit der Einführung der latenten Strukturanalyse zahlreiche Autoren bemüht waren, dem Anwender Algorithmen

bzw. fertige Programme zur Verfügung zu stellen. Eine Auswahl sei hier zitiert.

Für die Parameterschätzung bei der latenten Klassenanalyse entwickelten Lazarsfeld & Henry (1968, S. 74) ein Programm LASY, das auf einem algebraischen Verfahren beruhte. Von Henry (1966) stammt das Programm BAN, das unter den Namen NEWBAN geführt wird und Maximum-Likelihood-Schätzungen liefert. Von Cassady, Miller & Dingman (1968) stammt ein Algorithmus, der eine Verbesserung des von Madansky (1959) vorgeschlagenen Algorithmus zur Aufteilungsmethode darstellt.

Für die latente Klassenanalyse mit polytomen Items gibt Goodman (1974a) einen Algorithmus für Maximum-Likelihood-Schätzungen an. Eine Verallgemeinerung findet man bei Haberman (1977).

Für die latente Profilanalyse schlug Green (1970) einen Algorithmus basierend auf der Faktorisierungsmethode vor. Dieser wurde in dem Programm LPA von Mardberg (1971, 1975) und Isaacson (1972) verwendet. Von Takane (1976) wurde das Maximum-Likelihood-Schätzungen liefernde Programm MAXLPM entwickelt.

### **5.1.8 Probleme beim Schätzen**

In Fielding (1977) werden eine Reihe von Problemen diskutiert, die beim Schätzen von latenten Parametern auftreten.

Zwei Hauptfragen sind das Vorhandensein einer nicht allzu zeitaufwendigen Berechnungstechnik und die Konvergenz von Iterationsmethoden, falls solche verwendet werden. Z.B. ist die Konvergenz des von Takane (1976) vorgeschlagenen Algorithmus theoretisch noch nicht untersucht und offensichtlich nur für geeignete Ausgangswerte gegeben. Konvergenzuntersuchungen derartiger Algorithmen findet man in Haberman (1977). Nach Anderson (1959) haben Simulationsstudien und theoretische Überlegungen gezeigt, daß zumindest bis zu Stichprobenumfängen von 1000 sich sehr ungenaue Schätzungen ergeben können. Auch die Faktorisierungsmethode, obwohl mehr Information ausnutzend und in der Version von Green (1970) eine Rotation vermeidend, konvergiert meist sehr langsam. Außerdem liefert sie Schätzungen mit großen Varianzen, d.h. ungenaue Schätzungen. Dieses kann dazu führen, daß Modelle mit ganz unterschiedlichen Parameterstrukturen zu dem gleichen Datensatz passen.

Auch für Maximum-Likelihood-Schätzungen benötigt man hinreichend große Stichprobenumfänge bzw. hinreichend große Besetzungszahlen für die Kontingenztafeln der Antwortmuster. Lazarsfeld & Henry (1968, S. 82) berichten,

daß bei Simulationsstudien mit dem Programm BAN die Folge der Schätzer nicht immer konvergierte. Insbesondere war dieses nicht der Fall, falls die latenten Itemparameter nahe bei 0 oder 1 lagen. Auch hier zeigten sich große Varianzen der Schätzungen.

## 5.2 Signifikanztests

Die meisten Signifikanztests, die in der latenten Strukturanalyse vorgeschlagen worden sind, beziehen sich auf das Anpassungsproblem, d.h. auf die Frage, ob das angenommene Modell mit den Daten verträglich ist. Dieses wird gelegentlich reduziert auf das Problem, ob die angenommene Anzahl der latenten Klassen ausreichend ist. Man muß sich darüber im klaren sein, daß die angegebenen Tests immer nur Tests bezüglich der Nullhypothese sind, daß sich die manifesten Parameter aufgrund des Axioms der lokalen Unabhängigkeit als Summen von Produkten der latenten Parameter darstellen lassen. Selbst wenn diese Nullhypothese nicht verworfen werden kann, ist es möglich, daß andere Modelle die Daten noch besser erklären könnten. Ein zusätzliches Kriterium zum Verwerfen eines Modells sind sicher latente Parameter, die nicht zwischen 0 und 1 liegen.

Für die latente Klassenanalyse schlägt McHugh (1956) einen Chiquadratanpassungstest vor (vgl. auch Fielding, 1977). Für die latente Klassenanalyse mit polytomous Items werden in Goodman (1974b) Chiquadratanpassungstests und in Goodman (1974a) Likelihoodquotiententests angegeben. Beide Möglichkeiten werden auch in Haberman (1977) und Formann (1978b) diskutiert. Für das latente Distanzmodell geben Lazarsfeld & Henry (1968, S. 148) einen Test an. In Gibson (1962) findet man einen Chiquadratanpassungstest für das Problem, ob zusätzliche Items zu einem aufgrund von anderen Items aufgestellten Modell passen.

Für die latente Profilanalyse gibt Takane (1976) einen Likelihoodquotiententest für die Anpassungshypothese an. Darauf aufbauend wird ein Test diskutiert, der die Nullhypothese, daß  $q$  Klassen vorliegen, gegen die Alternativhypothese testet, daß mindestens  $q + 1$  Klassen vorliegen. Schließlich werden Tests, die spezielle Hypothesen bezüglich der Parameter betreffen, diskutiert.

## 6. Schätzung der latenten Variablen

Als Schätzung der latenten Variablen wird von Anderson (1959) der Rückschluß von dem vorgefundenen Antwortmuster eines Individuums auf seine Position im latenten Raum, speziell etwa seine Zuordnung zu einer bestimmten latenten Klasse, bezeichnet. In Torgerson (1958, S. 373) wird dieses als

Zuordnung von Skalenwerten an die Individuen, hier speziell für das lineare Polynommodell, diskutiert. Für die latente Klassenanalyse wird das Problem in Lazarsfeld & Henry (1968, S. 36-39, S. 68-70) behandelt. Spezialfälle, die die Skalierung von Antwortmustern betreffen, diskutiert Goodman (1975, 1978, 1979a). Für das latente Distanzmodell vergleiche man Lazarsfeld & Henry (1968, S. 138-139) und Torgerson (1958, S. 385), für das latente Profilmmodell Lazarsfeld & Henry (1968, S. 238-239) und Gibson (1962b), für das latente Klassenmodell mit polytomen Items Formann (1978b).

Die Zuweisung von Individuen zu Positionen des latenten Raumes entspricht der Durchführung einer Cluster-Analyse (Fielding, 1977). In Skene (1978) wird versucht, diesen Aspekt der latenten Strukturanalyse auch auf Probleme der Diskriminanzanalyse anzuwenden für den Fall von fehlenden Daten. Hingegen verwendet Ikuzawa (1968) eine Diskriminanzanalyse, um Individuen im latenten Klassenmodell zu identifizieren.

## **7. Vergleich mit anderen Verfahren**

### 7.1 Vergleich mit der Faktorenanalyse

Seitdem die latente Strukturanalyse eingeführt wurde, ist sie immer wieder mit der schon vorher existierenden in Konkurrenz stehenden Faktorenanalyse verglichen worden, und es ist versucht worden, die Zusammenhänge zwischen diesen beiden Verfahren herauszuarbeiten. In Green (1952) wird darauf hingewiesen, daß der Hauptunterschied der beiden Verfahren zu sein scheint, daß in der Faktorenanalyse nur die Interaktionen 1. Ordnung berücksichtigt werden, in der latenten Strukturanalyse aber auch Interaktionen höherer Ordnung. Jedoch glaubt Gebert (1978) festgestellt zu haben, daß in den Beispielen, die in der Literatur zur latenten Klassenanalyse publiziert wurden, die Interaktionen höherer Ordnung alle exakt oder nahezu Null sind und daß die latente Klassenanalyse anscheinend nur in solchen Fällen sinnvolle Interpretationen liefert.

Wie schon aus Green (1952) geschlossen werden kann, besteht die Möglichkeit, die Faktorenanalyse als Spezialfall in die latente Strukturanalyse einzubetten. Ähnlich finden Lazarsfeld & Henry (1968, S. 240-244) bei der Betrachtung stetiger latenter Räume eine Äquivalenz mit einem Einfaktormodell. In Gibson (1962a) wird auf die enge Beziehung der latenten Klassenanalyse zu dem faktorenanalytischen Modell von Dwyer (1937) hingewiesen, während Gibson (1956) die Identität des latenten Strukturmodells von Lazarsfeld (1950a) und des Konzepts der Proportionalprofile in der Faktorenanalyse bei Cattell (1944) und Cattell & Cattell (1955) nachweist. Hierhin gehört auch die Arbeit von Gibson (1959), in der als Vorteil der latenten Klassenanalyse gegenüber der Faktorenanalyse angeführt wird, daß man auf diese Weise die Probleme der

Schätzung der Kommunalitäten, der Rotation und der Nichtlinearität vermeiden könne. Aus dieser Zeit ist auch der Vergleich der beiden Verfahren durch Lazarsfeld (1959). Einen empirischen Vergleich führten Miller, Eyman & Dingman (1961) durch. Eine Anwendung beider Verfahren bei der Klassifikation geistig zurückgebliebener Patienten ergab eine bessere Übereinstimmung mit der üblichen Typologie für die latente Klassenanalyse.

In Fielding (1977) wird ebenfalls die Einbettung der Faktorenanalyse in das Modell der latenten Strukturanalyse diskutiert. Die Auffassung, daß es sich um zwei verschiedene Ansätze handelt, wird mit der praktischen Anwendung der latenten Strukturanalyse in Verbindung gebracht, bei der meist nur von einer latenten Variablen, die zudem noch klassiert ist, ausgegangen wird. Dieses führt nach Mardberg (1973) dazu, daß man von Faktorenanalyse spricht, wenn es um die Art des Zusammenhangs der manifesten Variablen geht, während latente Klassen eine Möglichkeit bieten, Individuen zu klassifizieren.

In Anderson (1959) wird ein allgemeines Skalierungsmodell aufgestellt, als dessen Spezialfälle sich Faktorenanalyse und latente Strukturanalyse ergeben. Daß in dieser Arbeit die Faktorenanalyse nicht als Spezialfall der latenten Strukturanalyse angesehen wird, liegt an einer eingeschränkten Definition der latenten Strukturanalyse, bei der die manifesten Variablen speziell dichotom sein müssen. Dieses ist wohl auch ein Grund dafür, daß in Takane (1972) das latente Klassenmodell als Spezialfall der Faktorenanalyse angesehen wird.

Wie in Lord & Novick (1968, S. 545) angemerkt wird, kann man bei der Faktorenanalyse die Annahme der lokalen Unabhängigkeit abschwächen zu der Annahme der linearen lokalen Unabhängigkeit. Dieses bedeutet aber im Gegensatz zu der Meinung von Takane (1972) nicht eine größere Allgemeinheit der faktorenanalytischen Modelle, sondern ist eine Folge der Nichtberücksichtigung höherer Interaktionen in der Faktorenanalyse. Wie Harper (1972) gezeigt hat, läßt sich die lokale Unabhängigkeit im latenten Klassenmodell dadurch abschwächen, daß man paarweise Abhängigkeit von Items zuläßt.

## 7.2 Vergleich mit der Guttman-Skalierung

Die ersten Arbeiten von Lazarsfeld (1950a, 1950b) erschienen in enger Nachbarschaft mit Arbeiten zu Skalierungsverfahren, insbesondere der Guttman-Skalierung. Schon McCarthy (1951) erkannte, daß man die Guttman-Skalierung als deterministischen Grenzfall der latenten Klassenanalyse ansehen kann. In Hays & Borgatta (1954) wird dann die latente Distanzanalyse direkt als ein probabilistisches Modell für die nichtperfekte Guttman-Skala verwendet. Die Notwendigkeit für ein solches Modell wird in den Schwierigkeiten gesehen, die nichtperfekte Skalen bei Anwendung der Guttmantheorie machen.

## 8. Anwendungen

Die latente Strukturanalyse ist auf die unterschiedlichsten Fragestellungen angewandt worden. Eine Auswahl sei hier aufgeführt.

Die latente Klassenanalyse wurde von Gill (1976) zur Klassifikation von Sehbehinderten, von Wiggins (1973) zur Analyse von Wahleinstellungen und Verhaltensänderungen, von Kadushin (1966) zur Bildung von unterschiedlich motivierten Gruppen, die eine Psychotherapie wünschten (vgl. auch Lazarsfeld & Henry, 1968, S. 88-92), sowie von Lazarsfeld (1950b) zur Untersuchung der Einstellung eingezogener Rekruten (vgl. auch Lazarsfeld & Henry, 1968, S. 106-111) verwendet. In McHugh (1956) wird eine auf Schumacher, Maxson & Martinek zurückgehende Untersuchung über kreative Fähigkeit bei Maschinenkonstruktionen und in Gibson (1951) das Problem von Hörfunk-Programm-Präferenzen (vgl. auch Lazarsfeld & Henry, 1968, S. 113-120) mit Hilfe des latenten Klassenmodells analysiert. In Miller, Sabagh & Dingman (1962) wird dieses Modell im Rahmen der medizinischen Diagnostik zur Klassifikation auf der Basis unterschiedlicher Mortalität verwendet, während Miller, Eyman & Dingman (1961) es zur Klassifikation von geistig zurückgebliebenen Patienten benutzen. Schließlich versucht Baker (1962) mit Hilfe des Modells das Bibliotheksproblem zu lösen, insbesondere die Klassifikation von Dokumenten, und Green, Carmone & Wachspress (1976) verwenden es, um Kundenverhalten bezüglich eines neu eingerichteten Fernsprechdienstes zu untersuchen.

Das erweiterte latente Klassenmodell mit „polytomous“ Items benutzt Hutchinson (1977) für Rückschlüsse auf die latente Unfallgeschwindigkeit bei Verkehrsunfällen und Formann (1978b) zur Bildung von Klassen von Suizidenten unter Verwendung einer Untersuchung von Bayreuther.

Die latente Profilanalyse wird von Magnusson & Ekehammar (1975) zur Analyse von Angstprofilen, von Goddard (1971) zur Untersuchung des Kontaktverhaltens von Firmenpersonal und von Mardberg (1972, 1973) zur Selektion und Klassifikation von Personal in der Industriepsychologie verwandt. Von Jonassen erhobenes Material über Charakteristiken von Gemeinden wird in Lazarsfeld & Henry (1968, S. 233-235) analysiert.

Das latente Distanzmodell verwenden Stouffer & Toby (1951), um Rollenkonflikte zu analysieren (vgl. Lazarsfeld & Henry, 1968, S. 142-148), während das latente Inhaltsmodell von Somers (1961) bei der Messung sozialer Rigidität verwendet wird (vgl. Lazarsfeld & Henry, 1968, S. 185-191). In Katz & Procter (1959) werden Soziogramme von Schulkindern mit Hilfe eines Markovkettenansatzes analysiert. Eine Reanalyse dieser Daten mit Hilfe des latenten Markovkettenmodells führen Lazarsfeld & Henry (1968, S. 258-263) durch. Ein solches Modell wendet van der Sanden (1978) auf Spiele des Gefangen-Dilemma-Typs an.

## Literatur

- Anderson, T. W. 1954. On estimation of parameters in latent structure analysis. *Psychometrika*, 19, 1-10.
- Anderson, T. W. 1959. Some scaling methods and estimation procedures in the latent class models. In U. Grenander (Hrsg.). *Probability and statistics*. New York: Wiley. 9-38.
- Baker, F. 1962. Information retrieval based on latent class analysis. *Journal of the Association for Computing Machinery*, 9, 512-521.
- Batchelder, W. H. & Narens, L. 1977. A critical examination of the analysis of dichotomous data. *Philosophy of Science*, 44, 113-135.
- Benini, R. 1928. Gruppi chiusi e gruppi aperti in alcuni fatti collettivi di combinazioni. *International Statistical Institute. Bulletin*, 23, 362-383.
- Brunk, H. D. 1958. On the estimation of parameters restricted by inequalities. *Annals of Mathematical Statistics*, 29, 437-454.
- Cassady, J. M., Miller, C. R. & Dingman, H. F. 1968. Latent class analysis: A direct approach. *Proceedings, 76th Annual APA Convention*, 209-210.
- Cattell, R. B. 1944. „Parallel proportional profiles“ and other principles for determining the choice of factors by rotation. *Psychometrika*, 9, 267-283.
- Cattell, R. B. & Cattell, A. K. S. 1955. Factor rotation for proportional profiles: analytical Solution and an example. *British Journal of Statistical Psychology*, 8, 83-92.
- Coleman, J. S. 1964. Models of change and response uncertainty. Englewood Cliffs, N. J.: Prentice-Hall.
- Cournot, A. A. 1838. Mémoire sur les applications du calcul des chances à la statistique judiciaire. *Journal de mathématiques pures et appliquées*, 3, 257-334.
- Dwyer, P. 1937. The determination of the factor loadings of a given test from the known factor loadings of other tests. *Psychometrika*, 2, 173-178.
- Fielding, A. 1977. Latent structure models. In: C. A. O'Muircheartaigh & C. Payne (Hrsg.). *Exploring data structures*. New York: Wiley. 125-157.
- Fischer, G. H. 1974. Einführung in die Theorie psychologischer Tests. Stuttgart/Wien: Huber.
- Formann, A. K. 1978a. A note on parameter estimation for Lazarsfeld's latent class analysis. *Psychometrika*, 43, 123-126.
- Formann, A. K. 1978b. The latent class analysis of polytomous data. *Biometrical Journal*, 20, 755-771.
- Gebert, A. 1978. Assoziationsstrukturanalyse und latent-class-model. *Zeitschrift für experimentelle und angewandte Psychologie*, 25, 46-54.
- Gibson, W. A. 1951. Applications of the mathematics of multiple factor analysis to the problems of latent structure analysis. Unveröffentlichte Dissertation. University of Chicago.

- Gibson, W. A. 1955. An extension of Anderson's solution for the latent structure equations. *Psychometrika*, 20, 69-73.
- Gibson, W. A. 1956. Proportional profiles and latent structure. *Psychometrika*, 21, 135-144.
- Gibson, W. A. 1959. Three multivariate models: factor analysis, latent structure analysis and latent profile analysis. *Psychometrika*, 24, 229-252.
- Gibson, W. A. 1962a. Extending latent class solutions to other variables. *Psychometrika*, 27, 73-81.
- Gibson, W. A. 1962b. Class assignment in the latent profile model. *Journal of Applied Psychology*, 46, 399-400.
- Gill, R. D. 1976. The model of latent structure analysis. *Statistica Neerlandica*, 30, **143-149**.
- Gilula, Z. 1979. Singular value decomposition of probability matrices: Probabilistic aspects of latent dichotomous variables. *Biometrika*, 66, 339-344.
- Goddard, J. B. 1971. Office communications patterns in central London. In: South east economic planning council: Office linkages in central London. Vol. II.
- Good, I. J. 1969. Some applications of the singular decomposition of a matrix. *Tech-nometrics*, 11, 823-833.
- Goodman, L. A. 1974a. Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61, 215-231.
- Goodman, L. A. 1974b. The analysis of systems of qualitative variables when some of the variables are unobservable. Part I - a modified latent structure approach. *American Journal of Sociology*, 79, 1179-1259.
- Goodman, L. A. 1975. A new model for scaling response patterns: An application of the quasi-independence concept. *Journal of the American Statistical Association*, 70, 755-768.
- Goodman, L. A. 1978. Analyzing qualitative/categorical data: Log-linear models and latent-structure analysis. London: Addison-Wesley.
- Goodman, L. A. 1979a. The analysis of qualitative variables using more parsimonious quasi-independence models, scaling models, and latent structures that fit the observed data. In: R. M. Merton, J. S. Coleman & P. H. Rossi (Hrsg.). Qualitative and quantitative social research: Papers in honor of Paul F. Lazarsfeld. New York: Free Press.
- Goodman, L. A. 1979b. On the estimation of Parameters in latent structure analysis. *Psychometrika*, 44, 123-128.
- Green, B. F. Jr. 1951. A general solution for the latent class model of latent structure analysis. *Psychometrika*, 16, 151-166.
- Green, B. F. Jr. 1952. Latent structure analysis and its relation to factor analysis. *Journal of the American Statistical Association*, 47, 71-76.
- Green, B. F. Jr. 1970. Parameter estimation in latent class analysis. Department of Psychology. Manuscript. Baltimore: John Hopkins University.

- Green, P. E., Carmone, F. J. & Wachspress, D. P. 1976. Consumer Segmentation via latent class analysis. *Journal of Consumer Research*, 3, 170-174.
- Haberman, S. J. 1974. Log-linear models for frequency tables derived by indirect observation: Maximum likelihood equations. *The Annals of Statistics*, 2, **911-924**.
- Haberman, S. J. 1977. Product models for frequency tables involving indirect observation. *The Annals of Statistics*, 5, 1124-1147.
- Harper, D. 1972. Local dependence latent structure models. *Psychometrika*, 37, **53-59**.
- Hays, D. G. & Borgatta, E. F. 1954. An empirical comparison of restricted and general latent distance analysis. *Psychometrika*, 19, 271-279.
- Henry, N. W. 1966. BAN, a Fortran program. Cambridge: Computing Laboratory.
- Hutchinson, T. P. 1977. Latent structure models applied to the joint distribution of drivers' injuries in road accidents. *Statistica Neerlandica*, 31, 105-111.
- Ikuzawa, M. 1968. Individual identification in latent class analysis: Application of canonical analysis and discriminant analysis. *Shinrigakuhyoron*, 11, 171-198.
- Isaacson, A. 1972. Latent profile analysis, a brief description. Technical Report No. 54. Stockholm: Royal Institute of Technology.
- Kadushin, C. 1966. The friends and supporters of psychotherapy. *American Sociological Review*, 31, 786-802.
- Katz, L. & Procter, C. 1959. The concept of configuration of interpersonal relations in a group as a time dependent stochastic process. *Psychometrika*, 24, 317-327.
- Koopmans, T. C. 1949. Identification problems in economic model construction. *Econometrica*, 17, 125-138.
- Koopmans, T. C. 1951. Identification problems in latent structure analysis. Manuscript. Cowles Commission Discussion Paper: Statistics, No. 360.
- Koopmans, T. C. & Reiersol, O. 1950. The identification of structural characteristics. *Annals of Mathematical Statistics*, 21, 165-181.
- Kuliback, S. 1959. Information theory and statistics. New York: Wiley.
- Lazarsfeld, P. F. 1950a. The logical and mathematical foundations of latent structure analysis. In: S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. H. Star & J. A. Clausen (Hrsg.). *Measurement and prediction*. Princeton: Princeton University Press. Chapter 10.
- Lazarsfeld, P. F. 1950b. Some latent structures. In: S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. H. Star & J. A. Clausen (Hrsg.). *Measurement and prediction*. Princeton: Princeton University Press. Chapter 11.
- Lazarsfeld, P. F. 1954. A conceptual introduction to latent structure analysis. In: P. F. Lazarsfeld (Hrsg.). *Mathematical thinking in the social sciences*. New York: The Free Press. Chapter 7.
- Lazarsfeld, P. F. 1955. Recent developments in latent structure analysis. *Sociometry*, 18, 391-403.

- Lazarsfeld, P. F. 1959. Latent structure analysis. In: S. Koch (Hrsg.). Psychology: A study of a science. New York: McGraw-Hill. Vol. 3, 476-535.
- Lazarsfeld, P. F. 1960. Latent structure analysis and test theory. In: H. Gulliksen & S. Messick. Psychological scaling theory and applications. New York: Wiley. Chapter 8.
- Lazarsfeld, P. F. 1961. The algebra of dichotomous systems. In: H. Solomon (Hrsg.). Studies in item analysis and prediction. Stanford: Stanford University Press. **111-157.**
- Lazarsfeld, P. F. & Dudman, J. 1951. The general solution of the latent class case. In: P. F. Lazarsfeld (Hrsg.). The use of mathematical models in the measurement of attitudes. Paper 5. RAND Research Memorandum No. 455.
- Lazarsfeld, P. F. & Henry, N. W. 1968. Latent structure analysis. Boston: Houghton Mifflin.
- Lippe, P. von der. 1973. Zur Skalierung komplexer Variablen mit der „Latent Structure Analysis“. Allgemeines Statistisches Archiv, 57, 333-356.
- Lord, F. M. 1953. An application of confidence intervals and of maximum likelihood to the estimation of an examinee's ability. Psychometrika, 18, 57-77.
- Lord, F. M. & Novick, M. R. 1968. Statistical theories of mental scores. Reading, Massachusetts: Addison-Wesley.
- Madansky, A. 1958. Identification and estimation in latent class analysis. Unveröffentlichte Dissertation. University of Chicago.
- Madansky, A. 1959. Partitioning methods of latent class analysis. Report P 1644. Rand Corporation.
- Madansky, A. 1960. Determinantal methods in latent class analysis. Psychometrika, **25, 183-198.**
- Madansky, A. 1968. Latent structure. In: International Encyclopedia of the Social Sciences. New York: The Macmillan Company. Vol. 9, S. 33-38.
- Magnusson, D. & Ekehammar, B. 1975. Anxiety profiles based on both situational and response factors. Multivariate Behavioral Research, 10, 27-43.
- Mardberg, B. 1971. L. P. A., a preliminary program for the solution of latent profile analysis, according to Green's method. Manuscript. Stockholm: University of Stockholm.
- Mardberg, B. 1972. Clustering jobs to general requirement profiles. Stockholm: Swedish Council for Personnel Administration.
- Mardberg, B. 1973. A model for selection and classification in industrial psychology. Supplement 19. Psychological Laboratories. University of Stockholm.
- Mardberg, B. 1975. LPA 2: A FORTRAN V Computer program for Green's Solution of latent class analysis applied to latent profile analysis. Educational and Psychological Measurement, 35, 163-166.
- McCarthy, P. J. 1951. A special review of „The American Soldier, Vol. IV“. Psychometrika, 16, 247-269.

- McDonald, R. P. 1962. A note on the derivation of the general latent class model. *Psychometrika*, 27, 203-206.
- McHugh, R. B. 1956. Efficient estimation and local identification in latent class analysis. *Psychometrika*, 21, 331-347.
- McHugh, R. B. 1958. Note on „Efficient estimation and local identification in latent class analysis“. *Psychometrika*, 23, 273-274.
- Meo, G. de. 1934. Su di alcuni indici atti a misurare l'attrazione matrimoniale in classificazioni dicotome. *Accademia delle Scienze Fisiche e Matematiche*. Naples. Rendiconto, 73, 62-77.
- Meredith, W. 1965. Some results based on a general stochastic model for mental tests. *Psychometrika*, 30, 419-440.
- Miller, C. R., Eyman, R. K. & Dingman, H. F. 1961. Factor analysis, latent structure analysis and mental typology. *The British Journal of Statistical Psychology*, 14, 29-34.
- Miller, C. R., Sabagh, H. & Dingman, H. F. 1962. Latent class analysis and differential mortality. *Journal of the American Statistical Association*, 57, 430-438.
- Procter, C. H. 1970. A probabilistic formulation and statistical analysis of Guttman scaling. *Psychometrika*, 35, 73-78.
- Sanden, A. L. van der. 1978. Latent Markov chain analysis of a value conflict in prisoner's dilemma games. *The British Journal of Mathematical & Statistical Psychology*, 31, 126-143.
- Skene, A. M. 1978. Discrimination using latent structure models. Compstat 1978. Proceedings in computational statistics. Wien: Physica-Verlag. 199-204.
- Somers, R. H. 1961. Latent content model of latent structure. Unveröffentlichte Dissertation. Columbia University.
- Stouffer, S. A. & Toby, J. 1951. Role conflict and personality. *American Journal of Sociology*, 56, 395-406.
- Takane, Y. 1972. Estimation of the recruitment latent class by least squares methods. *Japanese Psychological Research*, 14, 87-102.
- Takane, Y. 1976. A statistical procedure for the latent profile model. *Japanese Psychological Research*, 18, 82-90.
- Thurstone, L. L. 1925. A method of scaling psychological and educational tests. *Journal of Educational Psychology*, 16, 433-451.
- Torgerson, W. S. 1958. Theory and methods of scaling. New York: Wiley.
- Tucker, L. R. 1952. A level of proficiency scale for a unidimensional skill. *American Psychologist*, 7, 408.
- Waerden, B. L. van der. 1971. *Mathematische Statistik*. Berlin: Springer.
- Weinberg, W. 1902. Beiträge zur Psychologie und Pathologie der Mehrlingsgeburten beim Menschen. *Pflüger's Archiv für die gesamte Physiologie des Menschen und der Tiere*, 88, 346-430.

- Wiggins, L. 1973. Latent probabilities for attitude and behaviour processes. Amsterdam: Elsevier.
- Wilcox, R. R. 1979a. An alternative interpretation of three stability models. *Educational and Psychological Measurement*, **39**, 311-315.
- Wilcox, R. R. 1979b. Achievement tests and latent structure models. *British Journal of Mathematical and Statistical Psychology*, **32**, 61-71.
- Wottawa, H. 1979. Grundlagen und Probleme von Dimensionen in der Psychologie. Meisenheim: Hain.

## 7. Kapitel

# Clusteranalyse

*Hartmut-A. Oldenbürger*

## 1. Zur Entwicklung der Literatur

(Aufgabenstellung der vorliegenden Arbeit)

Clusteranalytische Verfahren konstruieren Gruppierungen von Objekten aufgrund ihrer ‚Ähnlichkeit‘.

Obwohl Methoden zur Realisierung dieses Ziels schon um 1935 (Tryon, 1939) vorgeschlagen wurden, fanden sie (im Vergleich zur weit verbreiteten Anwendung der Faktorenanalyse) wenig Beachtung. Erst die Veröffentlichung der inzwischen klassischen Monographie von Sokal und Sneath (1963) und gleichzeitig die größere Verfügbarkeit leistungsfähiger Rechenanlagen gaben die Anstöße zur breiteren Methodenentwicklung und expandierendem Einsatz clusteranalytischer Verfahren zunächst in der systematischen Biologie (Taxonomie), später auch in anderen empirischen Wissenschaften.

Für die Psychologie hat der Aufsatz ‚Hierarchical Clustering Schemes‘ von Stephen C. Johnson (1967) den entscheidenden Impuls zur methodologischen Diskussion und Anwendung von Clusteranalysen gegeben (Blashfield, 1980). Sie treten nun häufig an die Stelle von Faktorenanalysen, die in der Forschungspraxis sehr oft - inkorrekt erweise - zur Generierung von Variablen-**gruppierungen** eingesetzt worden sind. (Die Angemessenheit dieser Vorgehensweise insbesondere für die Skalenkonstruktion mit Hilfe von Tests diskutiert Revelle (1978, 1979); ein vergleichendes Beispiel für die WISC-R gibt Silverstein, 1980.)

Inzwischen kann die Entwicklung des Umfangs der Literatur als explosionsartig bezeichnet werden: Während in anderen Bereichen die Verdoppelungszeit der Publikationszahlen bei zwölf bis fünfzehn Jahren angesetzt werden kann, liegt sie für die clusteranalytische Literatur z.Zt. bei drei bis vier Jahren (Blashfield/Aldenderfer, 1978). Dieser Sachverhalt zeitigt unerfreuliche Begleiterscheinungen: (a) es bilden sich Cliques von Anwendern der Cluster-

analyse, die den Einsatz nur bestimmter Verfahren tradieren und (b) die Ausbildung gegenstandsbereichsspezifischer Jargons behindert die interdisziplinäre Kommunikation (Soweit die Ergebnisse einer Benutzerbefragung und einer Analyse des ‚Science Citation Index‘ von Blashfield und Aldenderfer, 1978.).

Davon scheint zumindest nach 1975 auch die methodologisch orientierte clusteranalytische Literatur betroffen zu sein. Während die wichtigen Lehrbücher und Monographien von Anderberg (1973), Bock (1974), Duran/Odell (1974), Hartigan (1975) Jardine/Sibson (1971), Sneath/Sokal (1973) und Vogel (1975) zumindest in ihren bibliographischen Teilen die verfügbare Literatur umfassend aufzuarbeiten versuchten, erscheinen nach 1975 gehäuft in die Verfahren einführende Artikel (z.B. Kopp 1978a, b, c) und elementare Lehrbücher, die die Fortentwicklungen der clusteranalytischen Methoden und Überprüfungsansätze nahezu unberücksichtigt lassen (vgl. dazu Steinhausen/Langer, 1977, Opitz 1980). 1980 allerdings wurde der beachtenswerte Studentext ‚Clusteranalysen‘ von Thomas Eckes und Helmut Roßbach publiziert. Dieser schließt die bestehende Lücke, indem er die wichtigsten methodologischen Ansätze des Gebiets aufarbeitet. Er mußte deshalb auf Intention, Planung und Gestaltung dieser Arbeit Einfluß haben. Der vorliegende Artikel erhebt danach keinerlei monographischen Anspruch, er soll stattdessen

- (a) eine gestraffte Übersicht der wesentlichsten clusteranalytischen Problemstellungen und Verfahren (mit den zugehörigen Referenzen), sowie
- (b) Ergänzungen und Hinweise zu methodologischen Weiterentwicklungen von Ansätzen und grundlegenden Innovationen geben.

Auf eine ausführliche Darstellung der mathematischen Grundlagen und der Vorgehensweise (an Beispielen und Veranschaulichungen) wird im allgemeinen verzichtet werden. (Siehe dazu Eckes/Roßbach, 1980. - Auch die kombinatorischen Ausdrücke zur Angabe der Anzahlen möglicher Lösungen von clusteranalytischen Problemstellungen werden nicht mitgeteilt. Diese führen zumeist lediglich zu der Aussage, das gestellte Problem sei wegen der immensen Anzahl der Möglichkeiten nicht exakt lösbar (siehe dazu Duran/Odell, 1974 und verschiedene Passagen in Bock, 1974 und Flachsmeyer, 1972); die Existenz von ‚backtracking-‘ und ‚branch and bound-‘ Verfahren aber belegt die Vordergründigkeit dieses Arguments.)

Einerseits soll also dem unbelasteten Leser zum Einstieg eine Strukturierung des Bereichs gegeben werden (advanced organizer), andererseits soll der kenntnisreichere Leser Hinweise auf und Anstöße zu Weiterentwicklungen erhalten.

## 2. Zur Datenerhebung und Datenstruktur

Dieses Kapitel setzt sich nicht mit genuin clusteranalytischen Fragestellungen auseinander. Vor der Darstellung der Verfahrensklassen (in Kapitel 3) erfolgt zunächst eine ordnende Charakterisierung des Datenmaterials und seiner Erhebung. Dazu werden einige notwendige Begriffe eingeführt und mit Hinweisen auf für die Psychologie kritische (Interpretations-)Probleme verknüpft.

Technisch gesehen liegen am Ausgangspunkt einer clusteranalytischen Fragestellung Daten gewöhnlich in (mindestens) einem der folgenden Modi vor:

- (a) als  $n \times m$  - Matrix  $[x_{ij}]$ ; durch die Eintragungen/Zellen dieser Matrix wird jeder Beobachtungswiederholung  $i$  eine Ausprägung eines Merkmals  $j$  (numerisch oder nichtnumerisch) zugeordnet. (Beispiele für Beobachtungswiederholungen sind Gegenstände, Personen, Situationen etc.; Beispiele für Merkmale sind Attribute, nominale Klassen, Variablen etc.)  
(Speziell im Fall numerisch interpretierter Ausprägungen der Merkmale können die Beobachtungswiederholungen statistisch als Realisationen von Zufallsvariablen angesehen oder als Punkte in einem (nicht notwendig euklidischen) mehrdimensionalen Raum aufgefaßt werden. Diese Sichtweisen sind auch für die andere Modalität, nämlich der Merkmale, einsetzbar.) Die Matrix  $[x_{ij}]$  wird hier häufig einfach als ‚Datenmatrix‘ bezeichnet.
- (b) als  $r \times r$  - Matrix  $[p_{kl}]$ ; durch die Eintragungen dieser Matrix wird jedem Objektpaar  $(o_k, o_l)$  aus einer Objektmenge  $O = \{o_1, \dots, o_k, o_l, \dots, o_r\}$  eine reelle Zahl  $p(o_k, o_l)$  oder kurz  $p_{kl}$ , zugeordnet, die die ‚Proximity‘ der beiden Objekte angibt. Der allgemeine Begriff der ‚Proximity‘ steht für mehrere Klassen von Maßen, etwa für die ‚Ähnlichkeit‘ bzw. ‚Unähnlichkeit‘, die ‚Nähe‘ bzw. ‚Entfernung (Distanz)‘, den ‚Zusammenhang‘ oder die ‚Abhängigkeit‘, die ‚Kovariation‘ oder ‚Interaktion‘ von Objektpaaren. (Beispiele für Objekte sind (Zufalls-)Variablen, Gegenstände, Lebewesen, Personen, Ereignisse, Situationen, Gegenstände der Anschauung oder des Denkens (Perzepte, Konzepte) etc. - Beispiele für Proximity-Maße sind Distanzen, Kovarianzen, Korrelationen, Maße erklärter Varianz, Kontingenzkoeffizienten, Häufigkeiten gemeinsamen Auftretens (insbesondere bei Ereignissen), Produktwahrscheinlichkeiten, bedingte Wahrscheinlichkeiten, Verwechslungswahrscheinlichkeiten, subjektive Ähnlichkeitsurteile etc.) Die Matrix  $[p_{kl}]$  wird hier häufig einfach als ‚Proximitymatrix‘ bezeichnet.

Da clusteranalytische Verfahren Gruppierungen zusammengehöriger Objekte zumeist auf der Basis von Proximitymatrizen generieren, ist es für die theoretisch-inhaltliche Interpretation der Resultate von herausragender Bedeutung, auf welchem Wege die Proximitymaße gewonnen wurden.

- (a) direkt: Im Rahmen z.B. einer Konkretisationserhebung kognitiver Strukturen werden Probanden nach ihrem subjektiven Urteil zur (paarweisen) Ähnlichkeit oder Zusammengehörigkeit von Perzepten oder Konzepten befragt. (Übliche Erhebungsparadigmata: Ratio-Scaling, (vollständiger) Paarvergleich, Sortierverfahren etc. - In diesem Fall ist es unzureichend, das Ergebnis der Erhebung, die Proximities, lediglich durch Repräsentationsverfahren (MDS und/oder Clusteranalyse) zu beschreiben und zu interpretieren, vielmehr ist ebenso eine theoretische Bemühung um die Explikation des Urteilsprozesses als kritisches Problem anzusehen.) Theoretische Aspekte der Ähnlichkeitsmessung erörtern Eckes/Roßbach, 1980, S. 36f..
- (b) indirekt: Im Rahmen z.B. einer Fragebogenuntersuchung werden Probanden aufgefordert, zu verschiedenen Statements die jeweilige Ausprägung ihrer persönlichen Zustimmung bzw. Ablehnung zu skalieren. Ergebnis dieses Vorgehens ist eine  $n \times m$  - Datenmatrix  $[x_{ij}]$  (Probanden x Statements). Will man eine kompakte Beschreibung der Datenstruktur z.B. durch Gruppierung herstellen, so ist zunächst - auf dem Begründungshintergrund der Fragestellung - die Modalität zu wählen, deren Elemente (Personen bzw. Items) als Objekte einer Clusteranalyse dienen sollen. Ferner ist begründet ein geeignetes Maß zu wählen oder zu konstruieren, welches die Proximity der Objekte beschreiben soll. Dieses Maß ist für das Resultat des Repräsentationsverfahrens (z.B. Clusteranalyse) konstitutiv. - Für die theoretisch-inhaltliche Einordnung des Gesamtergebnisses ist hervorzuheben, daß bei diesem Vorgehen die Ausprägung der ‚Ähnlichkeiten‘ bzw. ‚Unähnlichkeiten‘ zwischen den Objektpaaren nicht direkt erhoben, sondern errechnet wurde. Dabei benutzt man implizit bestimmte Variationsquellen (z.B. für die Korrelation zwischen Items die interindividuelle Varianz). Dieser Sachverhalt ist bei der Interpretation der Resultate als kritisches Problem anzusehen. (So wäre es für das obige Beispiel fehlerhaft, die durch Verfahrenseinsatz konstruierte Gruppierung der Items als Urteilsstruktur eines mittleren, ‚idealen‘ Individuums auszuzeichnen.)

Das Problem der Überführung der in der Forschungspraxis häufig vorliegenden Datenmatrix  $[x_{ij}]$  in eine Proximitymatrix  $[p_{kl}]$  kann nicht durch generelle Empfehlung gelöst werden. Auch innerhalb der Klassen der für numerisch interpretierbare, nominale und gemischte Daten angemessenen bivariaten Zusammenhangs- und Entfernungsmaße, werden durch die vielfältigen Koeffizienten verschiedene Aspekte der Kovariation formalisiert bzw. gemessen. Es ist deshalb im jeweiligen empirischen Kontext zu explizieren, unter welcher Fragestellung bzw. welchem Ziel das spezifische Proximitymaß vor dem Einsatz eines Repräsentationsverfahrens gewählt wurde. Die häufig geübte Praxis der voraussetzungskritischen, automatisierten Verwendung von Standardmaßen (z.B. Korrelationskoeffizient) ohne exploratorische Bemühungen (z.B. zur Kontrolle möglicher Artefaktquellen) ist unangemessen.

Darstellungen der verschiedenen Zusammenhangs- und Entfernungsmaße finden sich u.a. bei Anderberg (1973, S. 72ff.), Bock (1974, S. 24ff.), Sneath/Sokal (1973, S. 114ff.), Steinhausen/Langer (1977, S. 51 ff.); die wohl ausführlichste Übersicht zu symmetrischen und unsymmetrischen Kovariationsmaßen speziell für binäre Variablen findet sich bei Kleiter/Petermann (1977, S. 43ff.).

Als spezielle Proximitymaße sind **Distanzfunktionen** auch für clusteranalytische Verfahren von besonderer Bedeutung. (Sie dienen oft als Kriterium zur Messung der Homogenität von Objekten innerhalb von Gruppierungen bzw. zur Messung der Heterogenität von Objekten, die verschiedenen Gruppierungen angehören.) Deshalb werden die Eigenschaften dieser Maße hier im Rahmen einer Definition aufgeführt:

Seien  $o_k, o_l, o_s$  Elemente von  $O$ . Eine Funktion  $d: O \times O \rightarrow \mathbb{R}$  heißt Distanzmaß, genau dann wenn für alle  $o_k, o_l, o_s \in O$  gilt:

- (D.1)  $d(o_k, o_l) \geq 0$  Positivität
- (D.2)  $d(o_k, o_l) = 0 \Leftrightarrow o_k = o_l$  Identität
- (D.3)  $d(o_k, o_l) = d(o_l, o_k)$  Symmetrie
- (D.4)  $d(o_k, o_s) \leq d(o_k, o_l) + d(o_l, o_s)$  Dreiecksungleichung

Das Paar  $(O, d)$  heißt metrischer Raum;  $d$  wird auch Metrik genannt.

Gilt für alle  $o_k, o_l, o_s \in O$  die folgende Verschärfung von (D.4):

$$(D.5) \quad d(o_k, o_s) \leq \max \{d(o_k, o_l), d(o_l, o_s)\}$$

so heißt  $d$  **Ultrametrik**. (Anschaulich besagt (D.5), daß alle Tripel von Objekten gleichschenklige Dreiecke aufspannen, deren Basis kleiner ist als die gleichlangen Schenkel.) Johnson (1967) hat auf die besondere Bedeutung der Ultrametrik für einige Methoden der hierarchischen Clusteranalyse (siehe 3.3) hingewiesen: Diese können als Transformationsverfahren von empirischen **Unähnlichkeitsmaßen** (Geltung von mindestens (D.1) bis (D.3)) in Ultrametriken angesehen werden.

Da ultrametrische Distanzmaße bei Vorliegen einer Datenmatrix  $[x_{ij}]$  i. a. nicht direkt berechnet werden können, aber eine Reihe von Proximitymaßen die Distanzaxiome (D.1) bis (D.4) erfüllen oder metrische Äquivalente besitzen, erscheint es sinnvoll - soweit es von der Fragestellung her angemessen ist - zur Transformation von Datenmatrizen  $[x_{ij}]$  in Proximitymatrizen  $[p_{kl}]$  möglichst Metriken zu verwenden.

Bei Vorliegen numerisch interpretierbarer Eintragungen der Datenmatrix stellt die folgende Funktion eines der prominentesten Beispiele für eine Metrik dar:

$$d(o_k, o_l) = \left( \sum_j |x_{kj} - x_{lj}|^r \right)^{1/r}; r \geq 1$$

Diese Distanzfunktion heißt allgemein  $L_r$ -Metrik (in der Literatur zur mehrdimensionalen Skalierung: Minkowski- $r$ -Metrik). Wird der Exponent  $r$  spezifiziert, so entstehen als ausgezeichnete Fälle:

- für  $r = 1$ , die City-Block-Metrik
- für  $r = 2$ , die euklidische Metrik
- für  $r = m$ , die Supremum-Metrik

(Zu weiteren Eigenschaften dieser Abstandsfunktionen siehe Eckes/Roßbach 1980, S. 42ff. - Im originären Zusammenhang mit der mehrdimensionalen Skalierung stehen die Erörterungen von Ahrens 1974, z.B. S. 190ff.) Weitere Entfernungsmße, die z.B. auch bei Vorliegen binärer oder nominaler Merkmalsausprägungen eingesetzt werden können, finden sich in der zu den Proximitymaßen genannten Literatur.

Damit sind die wichtigsten Begriffe und Überlegungen zur Charakterisierung der Ausgangslage beim Einsatz clusteranalytischer Verfahren eingeführt. Im nächsten Kapitel wird das Vorliegen von Daten- und/oder Proximitymatrizen vorausgesetzt.

### 3. Problemstellungen und Verfahren

#### 3.1 Untermengenauswahl

Bemerkenswerterweise tritt das Problem der Selektion von Elementen aus einer gegebenen Menge unter Optimierung eines Kriteriums in verschiedenen methodischen und inhaltlichen Bereichen häufig auf, es wurde aber m. W. in der clusteranalytischen Literatur bisher nicht als eigenständige Fragestellung identifiziert. Deshalb seien einige Beispiele und Hinweise gegeben, die die Problemstellung charakterisieren und umreißen, sowie mögliche Lösungen andeuten:

- (a) Aus einer Menge von Prädiktoren mit dem Umfang  $n$  soll eine  $k$ -elementige (echte) Untermenge selektiert werden, so daß eine Kriteriumsvariable maximal prädiziert werden kann (Optimierungskriterium: z. B. adjustierter  $R^2$ ). Dies ist das Standardproblem der multiplen Regressionsanalyse.
- (b) Aus einer Menge von Variablen soll eine Untermenge vom Umfang  $k$  ausgewählt werden, so daß die Varianz aller restlichen Variablen maximal erklärt werden kann. Dieses Verfahren identifiziert die Stellvertreter eines Variablenkollektivs; es wird Interdependenzanalyse genannt (Boyce/Farhi/Weischedel, 1974, S. 67ff.) und kann als attraktive Alternative zur Faktorenanalyse angesehen werden. (Zwar entfällt hier das Problem der Faktorenterpretation, es wird aber ein gewisser Verlust an Varianzerklärung in Kauf genommen.)

- (c) Aus einer Menge von Gegenständen mit gegebenen Eigenschaften sollen Repräsentanten für die verbleibenden Objekte ausgewählt werden. Während die Interdependenzanalyse Merkmale (Variable) selegiert, soll nun nach zentralen Merkmalsträgern (Wechsel der Modalität) gesucht werden. Handelt es sich bei den Gegenständen z.B. um Perzepte oder Konzepte, so ist dies die Frage nach den Prototypen (z.B. von Kategorien), auf deren ausgezeichnete Rolle insbesondere Rasch (1975, 1978) hingewiesen hat.
- (d) Aus einer Menge von Gegenständen mit (numerischen) Merkmalen soll eine Untergruppe selegiert werden, deren Homogenität gegenüber dem Rest besonders groß ist. Diese Separation von Clustern vom Hintergrund (Yahil/Brown, 1976) kann in verschiedensten Anwendungsbereichen in Frage stehen, z.B. in der Astronomie bei der Herauslösung von Galaxien aus diffusen Verteilungen von Sternen, in der Medizin bei der Diagnostik von Auffälligkeiten in Röntgenbildern (Block et al., 1974) oder z.B. in der Psychologie bei der Identifikation besonders merkmalshomogener Personen in einer Gesamtgruppe. (Weitere Stichworte: Mustererkennung, Kategorienbildung, Klassifikationsprozesse.)
- (e) In engstem Zusammenhang mit der im vorigen Punkt angesprochenen Problemlage steht die Fragestellung der Analyse von Mischverteilungen. (Dieser Ansatz könnte auch unter Abschnitt 3.2 eingeordnet werden.) Man geht dabei von der Vorstellung aus, ein vorliegendes Sample sei aus einer Population gezogen, die aus mindestens zwei Subpopulationen mit Zufallsvariablen unterschiedlicher Mittelwertsvektoren und/oder Kovarianzmatrizen besteht. Aufgabe der Analyse ist die Bestimmung der Anzahl von Subpopulationen, die Schätzung der Parameter der multivariaten Zufallsvariablen und die Zuordnung der Beobachtungswiederholungen (Realisationen) zu den Subpopulationen. (Siehe Bock, 1974, S. 250ff.; Hartigan, 1975, S. 113ff.; Eckes/Roßbach, 1980, S. 92f.; Wolfe, 1970, 1978.) Der abschreckenden Komplexität dieser Aufgabenstellung steht die Attraktivität einer möglichen Lösung gegenüber: für nahezu alle Verfahren der multivariaten Datenanalyse (Regressions-, Varianz-, Interdependenz-, Faktor-, Diskriminanzanalyse etc.) stünde ein Überprüfungsmedium der Homogenität der Stichprobe zur Verfügung.
- (f) Aus einer Menge von Objekten mit gegebener Proximitymatrix soll eine Untergruppe selegiert werden, deren Proximities einer spezifischen - algebraisch formulierten - Charakteristik, z.B. einem System von (Un-)Gleichungen, möglichst gut genügen. Beispiele für derartige Problemstellungen sind etwa: die Auswahl Guttman- bzw. Rasch-skalierbarer Items aus einem (vorläufigen) Test; die Auswahl von Variablen mit paarweise auffällig linearem Zusammenhang; die Auswahl von Gegenständen der Anschauung oder des Denkens mit symmetrischer Ähnlichkeit oder die

Auswahl von Objekten, deren Proximities durch eine Ultrametrik besonders gut dargestellt werden können (Hubert, 1980). (Da sich solche Vorgehensweisen einem Immunisierungsverdacht besonders stark aussetzen, ist für sie die Entwicklung und Durchführung angemessener Überprüfungsverfahren dringend angezeigt.)

Diese kurze Übersicht soll deutlich machen: Gemeinsam ist den Ansätzen die Aufgabenstellung der Untermengenselektion, sie unterscheiden sich aber gravierend in ihren Zielsetzungen, die durch mathematische Formulierung von Optimalitätskriterien, sog. Zielfunktionen, zu explizieren sind (Beale, 1970). Die Verschiedenartigkeit der Zielsetzungen, die sich in der Heterogenität der Optimalitätskriterien hinsichtlich ihrer mathematischen Eigenschaften wider-spiegelt, läßt es nicht erwarten, daß zur Lösung des Auswahlproblems eine allgemein anwendbare optimale Verfahrensvorschrift (Algorithmus) angebar ist (abgesehen von der Möglichkeit der Totalenumeration für uninteressant kleine Unter- bzw. Obermengen).

Für verschiedene Problemstellungen allerdings liegen in der methodologischen Literatur, von der Forschungspraxis m. W. bisher unbeachtet, effiziente Lösungsverfahren vor: Schon 1967 publizierten Beale, Kendall und Mann einen Algorithmus, der für die unter den Punkten (a) bis (c) angesprochenen Aufgaben optimale Lösungen liefert. (Demgegenüber sind z.B. die von den verbreiteten Programmpaketen BMDP und SPSS realisierten Vorgehensweisen und Resultate (!) bei der Variablenauswahl der multiplen Regression als suboptimal zu bezeichnen.) Die programmtechnische Umsetzung des Verfahrens leisteten Boyce, Farhi und Weischedel (1974). (Sie weisen z.B. auch die Überlegenheit des Beale/Kendall/Mann-Algorithmus gegenüber den Standardauswahlverfahren der multiplen Regression (forward inclusion, backward elimination) aus. - In das Programmpaket BMDP-77 (siehe Dixon/Brown, 1977, S. 418ff.) wurde ein Algorithmus von Furnival/Wilson (1974) eingesetzt; siehe auch McHenry (1978).)

Für die komplexe Aufgabenstellung der Analyse von Mischverteilungen sei ergänzend auf das Vorliegen einer (weiteren) programmsprachlichen Formulierung (Fortran-Code) eines Verfahrens bei Hartigan (1975, S. 125ff.) hingewiesen. Hubert (1980) diskutiert eine Strategie der Datenanalyse zur Identifikation von Untermengen, für deren Elemente die Proximities bestimmten algebraischen Eigenschaften möglichst gut gehorchen.

Die methodischen Vorgehensweisen und Lösungsansätze zur optimalen Untermengenauswahl wurden bisher kaum beachtet; sie warten auf Bewährungsprüfungen.

### 3.2 Mengenerzeugung

Als Partition bzw. Zerlegung einer Menge 0 bezeichnet man eine Menge(nfamilie)  $Z(0)$ , deren Elemente (nichtleere) Untermengen von 0 sind:  $G_p, G_q \in Z(0)$  und folgende Forderungen erfüllen:

$$(Z. 1) \quad p \neq q \Rightarrow G_p \cap G_q = \emptyset$$

$$(Z. 2) \quad G_1 \cup G_2 \cup \dots \cup G_p \cup \dots = \bigcup_p G_p = 0,$$

d.h. die in  $Z(0)$  als Elemente enthaltenen Mengen  $G_p, G_q$ , die hier Gruppierungen genannt werden, sind paarweise elementenfremd bzw. disjunkt und sie erschöpfen die Menge 0 bzw. decken diese vollständig ab.

Zwei (triviale) Zerlegungen einer Menge 0 sind als extreme Fälle besonders ausgezeichnet: die feinste Partition (engl. ,Splitter') enthält als Elemente alle einelementigen Untermengen von 0; die grösste Partition (engl. ,lumber') enthält als Element nur die Menge 0 selbst.

Als Teilklasse clusteranalytischer Verfahren hat die Partitionierung generell folgende Aufgabenstellung: Man konstruiere bzw. suche zu einer gegebenen Objektmenge eine Zerlegung mit Gruppierungen  $G_p$ , für deren Elemente gilt: Gehören zwei Objekte zu gleichen Gruppierungen, so sollen sie einander möglichst ähnlich sein (innerhalb - homogen), gehören sie zu verschiedenen Gruppierungen, so sollen sie untereinander möglichst unähnlich sein (zwischen - heterogen). Als Ergebnis der Verfahrensanwendung wird die Feststellung der Anzahl der Gruppierungen und die Zuordnung der Objekte zu den Gruppen erwartet.

Die große Vielfalt der Methoden der disjunkten Gruppierung unterscheidet sich hauptsächlich danach,

- auf welche Informationen (Datenmatrix mit numerisch oder nicht numerisch interpretierbaren Merkmalsausprägungen und/oder Proximitymatrix) zurückgegriffen wird bzw. werden kann,
- ob und wie ein Optimierungskriterium (Zielfunktion) mathematisch spezifiziert wurde und
- nach welchem verfahrenstechnischen Konstruktionsprinzip die Methoden während des Analyseprozesses die Zerlegungen generieren (Enumeration; Modi der Zuordnung von Einzelobjekten zu bestehenden Gruppen, des Austausches von Objekten zwischen den Gruppen sowie der Separation und/oder Fusion von Gruppen).

Von verbreitetem forschungspraktischen Interesse sind Anwendungen, in denen zu den Objekten nur ein numerisches **Merkmal** gegeben ist oder betrachtet

wird (univariate Klassifikation, z.B. Aufteilung einer Personengruppe nach Intelligenz). In solchen Situationen ist zwar eine Aufteilung nach Perzentilen (Median, Terzile, Quartile etc.) gängig, diese läßt aber die Verteilungsform der Variablen unberücksichtigt. So geht beispielsweise die Trennung am Median für nicht-symmetrische Verteilungen mit einem (vermeidbaren) Informationsverlust einher. Eine optimale Aufteilung nach festgelegter Gruppenzahl und nahezu beliebig wählbarem Optimierungskriterium (z.B. Varianz zwischen den Gruppen) ist aber praktisch immer exakt möglich. Dies leistet ein Algorithmus von W. D. Fisher (1958), dessen Einsatz z.B. zur Variablenkategorisierung vor Durchführung einer Konfigurationsfrequenzanalyse angezeigt ist. Eine programmsprachliche Formulierung (Fortran-Code) des Verfahrens teilen Anderberg (1973, S. 228ff.) und Hartigan (1975, S. 141) mit.

Exakte Partitionierung (z.B. ebenfalls mit dem Fisher-Algorithmus) ist auch möglich, wenn die Aufteilung zwar nach mehreren Variablen (simultan) vorgenommen werden soll, die Anzahl der möglichen Partitionen aber dadurch erheblich eingeschränkt ist, daß den Objekten hinsichtlich einer zusätzlichen Variablen eine natürliche Ordnung unterliegt oder unterstellt wird (z.B. im Fall einer Konstruktion von Stufen der Entwicklung hinsichtlich mehrerer kognitiver Merkmale, wobei die **,natürliche Reihenfolge'** nach der Variablen „Alter“ eingehalten werden soll).

Im generellen Anwendungsfall, bei Vorliegen **multivariater Daten und/oder von Proximitymatrizen**, ist die Angabe von Verfahren, die zu exakten Partitionierungen führen, also eine Optimierung des spezifizierten Kriteriums garantieren (Rao, 1971), nicht möglich. (Abgesehen von den seltenen Fällen, in denen wegen des geringen Umfangs der Objektmenge und/oder einer Festlegung auf sehr wenige Gruppierungen eine Totalenumeration der Partitionen durchführbar und sinnvoll ist, siehe Duran/Odell, 1974, S. 32ff.) Deshalb werden heuristische Methoden verwendet, die z.B. von Startpartitionen ausgehen und diese durch Austausch von Elementen zwischen den Gruppen und Verschmelzen bzw. Aufteilen von Gruppen hinsichtlich des Wertes einer Zielfunktion sukzessive verbessern. Systematische Darstellungen zu Konstruktionsmethoden von (lage-)homogenen Gruppierungen, die das Varianzkriterium, die Kriterien der Spur oder Determinante der Kovarianzmatrix oder das Kriterium der Abstandsquadratsumme (Büttner, 1975) optimieren, werden von Bock (1974, S. 104ff.), Hartigan (1975, S. 74ff.), Späth (1975, S. 34ff.) und Steinhausen/Langer (1977, S. 100ff.) gegeben. Eine theoretisch orientierte Übersicht im Rahmen der Theorie(n) der kombinatorischen Programmierung liefern Balas/Padberg 1975.

Ergänzend sei auf die folgenden Weiterentwicklungen verwiesen:

- Ein effizienter (branch-and-bound-) Algorithmus zur globalen Optimierung des Varianzkriteriums wird von Herden/Steinhausen (1979, mit Fortran-Programm) mitgeteilt.

- Da viele Methoden den Nachteil haben, zur Konstruktion von Zerlegungen mit relativ gleich großen Gruppen zu tendieren, ist ein Verfahrensvorschlag von Eyes (1977a, 1977b, 1978) von besonderem Interesse: Grundidee ist dabei die Umschreibung lokaler Dichtezentren mit Hyperkörpern (Quadern, Ellipsoiden), die in ihrer Größe variieren und deren Lage im Raum verschoben wird. - Der oben angesprochene Nachteil kann auch mit Hilfe eines skalierungsinvarianten Partitionsverfahrens von Späth/Müller (1979) vermieden werden.
- Da die meisten Verfahren auf euklidische Optimierungskriterien zurückgreifen, sind sie anfällig gegen die Verletzung der normalerweise nicht zutreffenden Voraussetzung der Normalverteilung und gegen Vorliegen von Ausreißern (Outlier). Dies kann durch den Einsatz von resisterenteren bzw. robusteren Methoden vermieden werden, die z.B.  $L_1$ -Kriterien optimieren (siehe Späth, 1976, mit Fortran-Programm).
- Überlegungen zu Partitionierungsverfahren bei Vorliegen großer Objektmengen ( $n > 100$ ) stellen Marsten (1975) und Späth (1977a mit Fortran-Programm) an.

Zur **Evaluation von Partitionierungsverfahren** (Rand, 1971) legte Milligan (1980) eine empirische Untersuchung von vier K-Means-Algorithmen vor. Es wurden Datensätze mit bekannter Gruppierung durch sechs Arten von Fehlern überlagert und die Reklassifizierungsleistung der Verfahren erhoben. Diese erwies sich als sehr gut, wenn den Verfahren angemessene Startpartitionen vorgegeben werden. (Der Autor schlägt die Verwendung des Ergebnisses einer hierarchischen Group-Average-Clusteranalyse vor.) Als Konsequenz der Untersuchung stellen Milligan/Sokol 1980 ein Programm zur Durchführung des Jancey-K-Means-Algorithmus zur Verfügung.

**Programmiersprachliche Formulierungen** (Fortran) von Zerlegungsalgorithmen werden von Anderberg (1973, S. 307ff.), Hartigan (1975, S. 82ff.), Späth (1975, S. 36ff.) und Steinhausen/Langer (1977, S. 109ff.) mitgeteilt. Auch das in manchen Rechenzentren implementierte Programmpaket Clustan von Wis hart gestattet die Durchführung von Partitionierungsmethoden (1978, S. 43 ff.). Blashfield/Aldenderfer (1978) diskutieren und evaluieren verschiedene Computerprogramme, die Mengenzerlegungen realisieren.

Für die **Anwendung** sei folgendes Vorgehen empfohlen: Für eine festgelegte Anzahl von Gruppierungen verwende man verschiedene Zerlegungsalgorithmen, soweit vom Verfahren verlangt auch jeweils mit verschiedenen Startpartitionen, und wähle die Aufteilung, welche ein spezifiziertes Kriterium optimiert. (Eine Auswahl nach inhaltlich-theoretischen Gesichtspunkten, z.B. dem sog. „Kriterium der Interpretierbarkeit“ ist nicht angezeigt, da diese zu meist nicht nachvollziehbar ist - für Untersuchungen zu kognitiven Struk-

turierungen besteht zudem die Gefahr der Subjekt-Objekt-Konfundierung; zumindest ist die Mitteilung der hinsichtlich einer explizierten Zielfunktion besten Lösung erforderlich.) Für diese Optimallösung inspiziere man möglichst die Residuen (z.B. bei Klassifikation von Beobachtungswiederholungen die Abweichungen von gruppenspezifischen Mittelwerten für die einzelnen Variablen). Falls gewünscht, wiederhole man dieses Vorgehen für verschiedene Anzahlen von Gruppierungen. In diesem Fall sollte der Verlauf der Werte des Optimalitätskriteriums mitgeteilt werden. Die Auswahl der geeigneten Anzahl von Elementen der Zerlegung erfolgt danach möglichst durch probabilistische oder statistische Evaluation.

Zur **probabilistischen Bewertung der Zerlegbarkeit** von Objektmengen entwickelte parallel zu den Arbeiten von Lingoes/Cooper (1971, siehe auch Lingoes, 1973, S. 176ff., mit Fortran-Programm) und Schultz/Hubert (1973) insbesondere Ling (1971, 1972, 1973) einen exakten graphentheoretischen Ansatz (Kriterium: Verbundenheit ungerichteter Zufallsgraphen.) (Ergänzend sei dazu auf Frank, 1978, und Tinhofer, 1980, verwiesen.) Später wurden Tafeln publiziert (Ling, 1975, Ling/Killough, 1976), die eine partielle Evaluation disjunktiver Gruppierung auf der Basis dieser Theorie gestatten. Nachteil des graphentheoretischen Ansatzes ist die Notwendigkeit der impliziten Verwendung von Schwellenwerten (Cuts) zur Dichotomisierung von Proximitymatrizen (Informationsreduktion). Eine gravierendere Einschränkung der Anwendbarkeit des graphentheoretischen Konnektivitätskriteriums ergibt sich aus der im allgemeinen nicht zutreffenden Annahme, die Ausprägungen der Proximities könnten unabhängig voneinander frei variieren. Dies ist in der Regel nicht der Fall, so sind z.B. Distanzmaße tripelweise wegen der Geltung der Dreiecksungleichung voneinander abhängig. (So argumentieren auch Hubert/Baker, 1977a.)

Deshalb verdient ein Vorschlag von Lee/MacQueen (1980) besondere Beachtung. Zur Prüfung der statistischen Nullhypothese verknüpfen diese Autoren einen K-Means-Algorithmus mit einem Randomization-Test (Edgington, 1969, 1980). Dabei wird der Wert des Optimalitätskriteriums für die empirische Datenmatrix errechnet und mit optimalen Zielfunktionswerten verglichen, die man durch Verfahrensanwendung auf ‚zufällige Datenmatrizen‘ erhält. (Diese werden z.B. durch zufällige Permutation der empirischen Variablenwerte über die Beobachtungswiederholungen generiert.) Obwohl es sehr rechenzeitintensiv ist, sollte dieses Vorgehen als das Verfahren der Wahl für die statistisch konfirmatorische Beurteilung der Zerlegbarkeit von Objektmengen angesehen werden, zumal es ohne weiteres auch auf Datenmatrizen mit nicht numerisch interpretierbaren, z.B. nominalen Merkmalen anwendbar ist.

(Auch Hartigan diskutiert 1975 und 1977 Signifikanztests für Resultate von Zerlegungen (siehe 1975, S. 97ff., S. 135ff.). Seine Überlegungen basieren z.T. auf der statistischen Theorie der (multivariaten) Varianzanalyse. - Lei-

der legen solche Darstellungen dem Anwender völlig inkorrekte Vorgehensweisen nahe, z.B. die Durchführung und ‚statistische Interpretation‘ einer Varianzanalyse nach dem Einsatz einer Partitionierungsprozedur. Davon ist abzusehen, da durch die Verwendung kriteriumsoptimierender Zerlegungsverfahren **der** substantielle Bestandteil der statistischen Nullhypothese zerstört wird, nämlich die zufällige Zuordnung zu den Gruppen. Spielarten dieses gravierenden Fehlers finden sich in der inhaltlichen, anwendungsorientierten Literatur sehr häufig, z.B. auch bei der ‚statistischen Beurteilung‘ der Resultate von Techniken der schrittweisen Regression. - Die oben angesprochene Prüfstrategie unter Verwendung von Randomization-Tests unterliegt dieser Kritik nicht, weil sie die Optimierungsverfahren in die Prozedur zur Generierung der Verteilung der statistischen Prüfgröße einbaut.)

Soll nach der empirischen Untersuchung die Übereinstimmung des Resultats von Zerlegungsverfahren mit gegebenen, z.B. theoretisch, a priori konstruierten Partitionen statistisch geprüft werden, so stehen die üblichen nicht-parametrischen Tests zu Kontingenztafeln und die zugehörigen Maße praktischer Signifikanz bzw. Bedeutsamkeit zur Verfügung. Dabei ist wegen der Gewichtungsmöglichkeit in diesem Zusammenhang insbesondere auf Cohens ‚Kappa‘ zu verweisen. (Siehe dazu auch Hubert, 1979.) - Verschiedene (Distanz-)Maße für den Vergleich von Partitionen diskutieren Boorman/Arabie, 1972 und Arabie/Boorman, 1973.

Abschließend bleibt festzustellen, daß die Ausführungen dieses Abschnitts sich bisher auf Vorgehensweisen bezogen, die die Konstruktion lagehomogener Gruppierungen von Objekten zum Ziel haben. Die Suche nach optimalen Mengenzerlegungen kann durch geeignete mathematische Formulierung der Bewertungsfunktion aber auch zur Realisierung anderer, komplexerer Zielvorstellungen eingesetzt werden. Nach dem Verweis auf die Analyse von Mischverteilungen (siehe Abschnitt 3.1, Punkt(e)) seien folgende **Ansätze** benannt:

- (a) Woodward/Bentler (1979) zerlegen die Menge der Items eines Tests zur Optimierung des Reliabilitätskoeffizienten  $\alpha$  in zwei elementenfremde Gruppen.

Ihren einfachen Austauschalgorithmus zur (binären) Zerlegung (Darstellung durch einen Vorzeichenvektor) bezeichnen sie als hoch effektiv. (Deshalb schlagen sie seine Verwendung auch für die Konstruktion lagehomogener Gruppierungen vor.) Von besonderem allgemeinen Interesse ist dazu ihre - eigentlich schon aus anderen Zusammenhängen bekannte - Argumentation: Wenn der Austauschalgorithmus (lediglich) mit einer Wahrscheinlichkeit von  $p$  (z.B.  $p = .2$ ) bei einer zufälligen Ausgangspartition zum globalen Optimum führt (Die Erreichung einer lokalen Optimalkonfiguration ist ohnedies garantiert.), so besitzt er bei  $N$ -maligem Aufruf (z.B.  $N=20$ ) mit verschiedenen zufälligen Startpartitionen eine Wahrscheinlichkeit von  $P = 1-(1-p)^N$  (z.B.  $P = 1-(1-.2)^{20}=.988$ ), daß sich

unter den lokal optimalen Endpartitionen die global beste Zerlegung befindet. (Diese interessante Argumentationsfigur gibt zu der Frage Anlaß, inwieweit es ökonomisch ist, Entwicklungs- und Rechenzeit in kriteriumsspezifische Optimierungsverfahren zu investieren, wenn mit einem Austauschalgorithmus eine generell anwendbare Prozedur vorliegt; allerdings ist vor deren Beantwortung eine jeweilige Abschätzung der Wahrscheinlichkeit  $p$  von nöten, für bestimmte Problemstellungen zum Optimum zu gelangen.)

- (b) Bock (1974, S. 195ff.) diskutiert die Charakterisierung von disjunkten Gruppen durch Hyperebenen. Dabei wird die Zuordnung von Objekten zu Klassen z.B. nicht nach ihrer Entfernung vom Gruppencentroid (bzgl. der numerischen Merkmale) vorgenommen bzw. beurteilt, sondern nach ihrer (senkrechten) Entfernung von einer gruppenspezifischen, im mehrdimensionalen Raum aufgespannten Fläche. Ein Beispiel für diese Problemstellung ist etwa die optimale Aufteilung von Personen in Gruppen, so daß der Variablenraum gruppenspezifisch möglichst einfach (d.h. mit niedriger Dimensionalität) durch Hauptkomponenten repräsentiert werden kann.
- (c) Im Anschluß an die unter (b) aufscheinende Grundidee wird hier eine Klasse von Verfahren vorgeschlagen, die mit der Bezeichnung ‚Klassifikationsansatz‘ überschrieben sein soll. Ihre gemeinsame Problemstellung sei wie folgt charakterisiert: Hinsichtlich einer Modalität (der Datenmatrix: Merkmale bzw. Merkmalsträger) sei eine Aufteilung so konstruiert, daß hinsichtlich der (jeweils) anderen Modalität - innerhalb der Gruppen - eine möglichst gute Repräsentation durch ein multivariates Analyseverfahren bzw. -modell erzielt wird.

Ein spezielles Beispiel wäre die unter (b) beschriebene Aufgabenstellung der ‚Klassifikations-Hauptkomponentenanalyse‘.

Ein weiteres Beispiel ist die ‚(multiple) Klassifikationsregression‘, die eine optimale Aufteilung der Beobachtungswiederholungen so sucht, daß innerhalb der Gruppierungen eine möglichst gute Varianzaufklärung einer Kriteriumsvariable durch einen Satz von Prädiktoren erreicht wird. (Dieses Verfahren ‚hinterfragt‘ den allgemeinen Anspruch der ‚Multiple Regression as a General Data Analytic System‘ (Cohen, 1968). - Den Spezialfall nur eines Prädiktors mit homogenen (identischen) Steigungskoeffizienten für die (nicht essentielle) Restriktion auf zwei Gruppen untersucht bereits Mustonen (1978).)

Weiterhin generiert dieser allgemeine, als Heuristik dienende Ansatz (auschöpfend) z.B. die Verfahren der ‚Klassifikationsanalyse kanonischer Korrelation‘, der ‚Klassifikations-Mengenerlegungsanalyse‘ und der ‚Klassifikationsanalyse hierarchischer Clusterung‘. Schließlich können sogar Verfahren der ‚Klassifikations-Varianzanalyse‘ und der ‚Klassifikations-Diskriminanzanalyse‘ einbezogen werden.

Offensichtlich steht der hier entwickelte allgemeine Klassifikationsansatz in engster Beziehung zur klassischen Analyse latenter Klassen (z.B. Lazarsfeld/Henry, 1968); er könnte in der Anwendung vornehmlich der Homogenitätsprüfung von Modellapproximationen zu Samples dienen; der methodologischen Literatur bleibt die Bereitstellung von Überprüfungsverfahren vorbehalten.

Für die praktische Umsetzung der hochkomplexen Aufgabenstellungen der Analyse von Mengenzerlegungen, z.B. des Klassifikationsansatzes, wird man zur Optimierung der (jeweils zu spezifizierenden) Zielfunktion vorläufig auf den Einsatz der relativ generellen Austauschverfahren zurückgreifen (siehe z.B. Banfield/Basill, 1977, mit Fortran-Programm) und die beste Lösung der Anwendung eines Transferalgorithmus auf verschiedene Startpartitionen auswählen. - Im Anwendungsfall sind exploratorische Bemühungen (sensu Tukey, 1977), z.B. in Form einer genauen Residueninspektion, und die Entwicklung und Durchführung statistischer Prüfungen, z.B. durch angemessen konstruierte Randomization-Tests, unverzichtbar, da sonst die Gefahr der Interpretation von Artefakten zu groß würde. Dabei ist der rechenzeitintensive Einsatz von Computern zur genauen Prüfung von Verfahrensresultaten bzw. Modellen nicht nur notwendig, sondern auch sinnvoll und wünschenswert.

### 3.3 Hierarchische Clusteranalysen

#### **3.3.1 Einordnung und Charakteristik**

Hierarchische Verfahren sind in der Psychologie die am weitesten verbreiteten Methoden aus dem Bereich der Clusteranalyse (Blashfield/Aldenderfer, 1978). Der Boom in der Methodendiskussion und der inhaltlich orientierten Anwendung geht auf den schon klassischen Artikel ‚Hierarchical Clustering Schemes‘ von Stephen C. Johnson (1967) zurück (Blashfield, 1980). Dabei hatte der Autor lediglich Verfahren der ‚numerischen Taxonomie‘ aus der Biologie übernommen (Sokal/Sneath, 1963) (und parallel zu Hartigan (1967) die Charakterisierbarkeit der Verfahrensergebnisse durch die ultrametrische Ungleichung (siehe Abschnitt 2) aufgezeigt). Die breite Aufnahme der für die Psychologie neuen Methodenklasse ist wohl auch weniger auf die allgemeine Angemessenheit der Verfahren zur strukturellen Repräsentation von Proximitymatrizen zurückzuführen, vielmehr traf Johnsons Brückenschlag auf einen Boden, der durch das stark aufkeimende ‚Unbehagen in der Faktorenanalyse‘ (Kallina, 1967) aufbereitet war und damit die Suche nach alternativen Analysemethoden förderte.

Ausgangssituation für den Einsatz von Verfahren der hierarchischen Clusteranalyse (HCA) ist das Vorliegen einer Proximitymatrix (von Ähnlichkeitsma-

ßen  $s_{kl}$  oder Unähnlichkeitsmaßen bzw. Distanzen  $d_{kl}$ ) zu einer bestimmten Objektmenge  $O$  mit  $n$  Elementen. Aufgabe von Verfahren der HCA ist die Repräsentation der Proximities durch die Angabe einer geordneten Folge von Zerlegungen  $Z(0)_0, Z(0)_1, \dots, Z(0)_h, \dots, Z(0)_{n-1}$ , wobei  $Z(0)_0$  die feinste und  $Z(O)$ , die größte Partition zu  $O$  ist. (Zugang sensu Ward, 1963; zur Definition der hier verwendeten Begriffe siehe Abschnitt 2 und 3.2.)

Ferner soll für alle  $h$  mit  $0 \leq h \leq n-2$  gelten, daß  $Z(0)_{h+1}$  eine strikte Vergrößerung von  $Z(0)_h$  ist, d.h. (genau) eine Gruppierung aus  $Z(0)_{h+1}$  ist durch Vereinigung von (mindestens) zwei Gruppierungen aus  $Z(0)_h$  entstanden. Umgekehrt heißt  $Z(0)_h$  auch Verfeinerung von  $Z(0)_{h+1}$ . Entnimmt man also aus beliebigen zwei Zerlegungen der geordneten Folge jeweils eine Gruppierung, so sind diese entweder elementenfremd oder eine Gruppierung ist Untermenge der anderen. Dieser Sachverhalt charakterisiert ein hierarchisches (Mengen)System, siehe z.B. (Bock, 1974, S. 360.)

Geordnete Folgen von Zerlegungen können explizit oder grafisch durch sog. Dendrogramme angegeben werden; Beispiele finden sich in der unten aufgeführten Literatur. Diese veranschaulichen die sukzessive Vergrößerung der Partitionen durch Fusion von Gruppierungen bzw. umgekehrt die sukzessive Verfeinerung durch Separation von Gruppierungen. Außerdem eröffnen sie die Möglichkeit, den Zusammenlegungen bzw. Trennungen bestimmte (numerische) Niveaus zuzuordnen; aus ihnen können für alle Paare von Objekten aus  $O$ , über die Höhe der Knoten, bei der die Objekte durch Kantenfolgen zusammengeführt werden, Proximities  $s_{kl}$  bzw.  $d_{kl}$  bestimmt werden. Ist man bei der Analyse von Unähnlichkeitsmaßen oder Distanzen ausgegangen, so erfüllen die  $d_{kl}$  die ultrametrische Ungleichung. (Zur Definition siehe Abschnitt 2; im Fall der Analyse von Ähnlichkeitsmaßen erfüllen die  $s_{kl}$  die der Ultrametrik äquivalente Eigenschaft: für alle  $o_k, o_l, o, EO$  gilt  $s_{ks} \leq \min\{s_{kl}, s_{ls}\}$ ; hier wird von einzelnen Verfahren abgesehen, die sog. Inversionen zulassen (siehe Eckes/Roßbach, 1980, S. 73), und deshalb verschiedentlich zu nicht interpretierbaren Resultaten führen (ein empirisches Beispiel findet man bei Steinhäusen/Langer, 1977, S. 92f.).)

Da Dendrogramme und Matrizen ultrametrischer Proximities einander eindeutig zugeordnet sind, können hierarchische Clusteranalysen als Transformationsverfahren von empirischen Proximitymatrizen in Ultrametriken charakterisiert werden. Dies gibt Anlaß zu einer wichtigen Reformulierung der Aufgabenstellung von HCA-Methoden: zu einer Matrix, z.B. von Unähnlichkeitsmaßen  $[d_{kl}]$ , sollen sie eine ultrametrische Matrix, z.B.  $[s_{kl}]$ , so konstruieren, daß hinsichtlich einer spezifizierten Zielfunktion  $Z = f([d_{kl}], [s_{kl}])$  die Ultrametrik eine optimale Repräsentation der (empirisch) gegebenen Proximitydaten darstellt. - Eine Liste solcher Optimalitätsmaße gibt Cormack (1973, S. 338); diese wurden zwar verschiedentlich zur nachträglichen Evaluation der Resultate von HCA-Verfahren vorgeschlagen, (so nennt Johnson (1967) z.B.

die Rangkorrelation), ihr konstruktiver Einsatz zur Generierung von Lösungen eines (diskreten) Optimierungsproblems wurde bisher jedoch allzu stark vernachlässigt. Die Methoden der HCA erhalten dann einen lediglich heuristischen Charakter, d.h. die Resultate der Verfahrensanwendungen führen zu verschiedenen Vorschlägen zur Lösung des Repräsentationsproblems. In der Forschungspraxis äußert sich dies manchmal in einer Verfahrensselektion nach persönlicher Vorliebe und in einer schwer nachvollziehbaren Auswahl ihrer Resultate nach einem Kriterium der ‚theoretischen Angemessenheit‘ bzw. ‚Interpretierbarkeit‘. (Ein Vorschlag für ein begründeteres Vorgehen wird in Abschnitt 3.3.4 dargestellt. - Es gibt aber auch Autoren, die den konstruktiven Optimierungsansatz verfolgen: Hartigan (1967), Kruskal/Carroll (1969), Carroll/Pruzansky (1980).)

Systematische übersichten und zahlreiche Beispiele zu den (heuristischen) Methoden der HCA geben: Anderberg (1973, S. 131ff.), Bock (1974, S. 356ff.), Eckes/Roßbach (1980, S. 63ff.), Hartigan (1975, S. 191ff.), Jardine/Sibson (1971, S. 45ff.), Opitz (1980, S. 96ff.), Späth (1975, S. 147ff.), Steinhausen/Langer (1977, S. 73 ff.) und Vogel (1975, S. 249ff.).

Zur Konstruktion von geordneten Folgen von Partitionen der Objektmenge gibt es zwei übliche Gruppen von Vorgehensweisen: Die agglomerativen Verfahren beginnen mit der feinsten Partition und legen sukzessive Gruppierungen nach einem Kriterium der maximalen Ähnlichkeit bzw. minimalen Distanz zusammen. Die subdivisiven Verfahren beginnen mit der gröbstens Partition und zerlegen sukzessive die Gruppierungen nach einem Kriterium der minimalen Ähnlichkeit bzw. maximalen Distanz. Es ist aber durchaus nicht zwingend, den Analyseprozeß bei den trivialen Partitionen zu beginnen, vielmehr ist es mindestens ebenso sinnvoll, zunächst die Konstruktion einer optimalen Mengenerlegung (siehe Abschnitt 3.2) mit einer mittleren Gruppenzahl, z.B. von ca.  $n/2$ , vorzunehmen, um dann, von dieser Partition ausgehend, die Folge der Zerlegungen agglomerativ ‚nach oben‘ und subdivisiv ‚nach unten‘ zu vervollständigen. Für die Praxis böte sich gerade ein solches Vorgehen an. Da sehr häufig nicht das Gesamtdendrogramm, sondern ohnedies eine Partition ‚mittleren Niveaus‘ zur Interpretation herangezogen wird, liegt es zumindest für diese Fälle nahe, die Bemühung um Optimierung (der Abbildungstreue) auch in diesem Bereich anzusetzen.

Da für alle gängigen HCA-Verfahren die Ausgangsdaten (Proximities) Ähnlichkeiten bzw. Distanzen zwischen einzelnen Objekten angeben, ist die definitorische Festlegung dessen, was unter Ähnlichkeit bzw. Distanz von Gruppierungen anzusehen ist, zentrales methodengenerierendes Prinzip. Betrachtet man z.B. Distanzen, so gilt es folgende Kriteriumsfunktion zu spezifizieren:

$$(F.1) \quad d(G_p, G_q) = \min_{\substack{\text{Funktion} \\ o_k \in G_p, o_l \in G_q}} \{d(o_k, o_l)\}$$

Setzt man z.B. fest, daß man agglomerativ vorgehen will, und daß für Methode  $M_1$ : ‚Funktion‘ ≡: ‚Min‘, für  $M_2$ : ‚Funktion‘ ≡: ‚Max‘ und für  $M_3$ : ‚Funktion‘ ≡: ‚arithmetisches Mittel‘, so hat man drei prominente HCA-Verfahren konstruiert:

- $M_1$  heißt ‚Single Linkage‘ oder ‚Minimum Methode‘; für den Fall der Fusion von  $G_p$  und  $G_q$  garantiert sie, daß zwei Objekte (je eines aus  $G_p$  und  $G_q$ ) mindestens eine Distanz von  $d(G_p, G_q)$  aufweisen.

- $M_2$  heißt ‚Complete Linkage‘ oder ‚Maximum Methode‘; sie garantiert, daß in der neuen Gruppierung  $G_p \cup G_q$  keine paarweise Distanz zwischen Objekten größer als  $d(G_p, G_q)$  ist, weshalb sie auch manchmal ‚Diameter Methode‘ genannt wird.

- $M_3$  heißt ‚Group-Average Methode‘.

Dieser Darstellungsmodus verdeutlicht, welche Vielfalt von heuristischen Verfahren mit einfachen Mitteln konstruierbar ist: es sind prinzipiell unendlich viele. Damit wird wiederum die Notwendigkeit des konsequenten konstruktiven Einsatzes von Optimierungsverfahren unterstrichen.

Andererseits kann im Zusammenhang mit (F.I) auch leicht erklärt werden, warum agglomerative Verfahren sich gegenüber subdivisiven Verfahren so großer Beliebtheit erfreuen: Sie machen nicht so viel Arbeit bzw. sind leicht von Hand durchführbar ( $n < 20$ ). Während bei fusionierendem Vorgehen z.B. auf unterster Stufe (feinste Partition) lediglich  $n(n-1)/2$  Werte  $d(G_p, G_q)$  (die alle noch gleich  $d(o_k, o_l)$  sind) untersucht werden müssen, sind es beim separierenden Vorgehen auf oberster Stufe (größte Partition)  $2^{n-1}-1$  mögliche (binäre) Zerlegungen mit zugehörigen Werten  $d(G_p, G_q)$ . Diese ‚Ökonomieüberlegung‘ spricht aber keineswegs von vornherein für die agglomerativen Verfahren, da diese auf niedriger Stufe aufgrund relativ weniger Werte durch die Zusammenlegungen Vorentscheidungen treffen, die in späteren Stadien des Analyseprozesses u.U. ‚ungünstige‘ Fusionen erzwingen. Man kann aber (wiederum, s. o.) davon ausgehen, daß Anwender häufig an der Interpretation relativ weniger großer Gruppen interessiert sind. Deshalb sollte deren Zusammensetzung auch möglichst sorgfältig erfolgen.

### **3.3.2 Agglomerative Verfahren**

Speziell für die kompakte tabellarische Darstellung der verbreiteten agglomerativen Prozeduren wird in der Literatur eine Schreibweise tradiert, die auf Lance/Williams (z.B. 1967) zurückgeht. Wurden auf einer bestimmten Stufe der Analyse zwei Gruppierungen  $G_p$  und  $G_q$  aufgrund z.B. des Kriteriums der minimalen Distanz bzw. der maximalen Ähnlichkeit für die Aggregation aus-

Tabelle: Agglomerative hierarchische Clusteranalysen.

Name	$\alpha_p$	$\alpha_q$	Parameter	Eigenschaften		
				$\beta$	$\gamma$	monoton invariant bei ultrametrisch monotoner Transf.
1 Single Linkage	1/2	1/2	0	-1/2	Ja	Ja
2 Complete Linkage	1/2	1/2	0	1/2	Ja	Ja
3 Group Average	$n_p/(n_p+n_q)$	$n_q/(n_p+n_q)$	0	0	Nein	Nein
4 Weighted Average	1/2	1/2	0	0	Ja	Nein
5 Centroid	$n_q/(n_p+n_q)$	$n_p/(n_p+n_q)$	$-n_p n_q / (n_p + n_q)^2$	0	Nein	Nein
6 Median	1/2	1/2	$-1/4$	0	Nein	Nein
7 Minimum Variance	$(n_p+n_r)/(n_p+n_q+n_r)$	$(n_q+n_r)/(n_p+n_q+n_r)$	$-n_r/(n_p+n_q+n_r)$	0	Ja	Nein
8 Flexible	$(1-\beta)/2$	$(1-\beta)/2$	$\beta(<1)$	0	Ja	Nein

Gebrauchliche alternative Bezeichnungen:

- Zu 1: Nearest Neighbor Method, Connectedness Method, Minimum Method
- Zu 2: Furthest Neighbor Method, Diameter Method, Maximum Method
- Zu 3: Unweighted Pair-Group Method Using Arithmetic Averages (UPGMA)
- Zu 4: Weighted Pair-Group Method Using Arithmetic Averages (WPGMA)
- Zu 5: Unweighted Pair-Group Centroid Method (UPGMC)
- Zu 6: Weighted Pair-Group Centroid Method (WPGMC)
- Zu 7: Error Sum of Squares Method (HGROUP)

gewählt, so ist zunächst zu bestimmen, welche Distanz die neue Gruppierung  $G_p \cup G_q$  zu den anderen Gruppen  $G_r$  haben soll. Durch Spezifikation der Parameter  $\alpha_p$ ,  $\alpha_q$ ,  $\beta$  und  $\gamma$  der folgenden Rekursionsformel ergeben sich dann verschiedene agglomerative HCA-Verfahren:

$$(F.2) \quad d(G_r, G_p \cup G_q) = \alpha_p d_{rp} + \alpha_q d_{rq} + \beta d_{pq} + \gamma |d_{rp} - d_{rq}|$$

Dabei wurde z.B.  $d(G_r, G_p) \equiv: d_{rp}$  verwendet.

In der Tabelle stehen  $n_p$ ,  $n_q$ , und  $n_r$  für die Anzahl der Elemente in den Gruppen  $G_p$ ,  $G_q$  und  $G_r$ . (Ähnliche Übersichten geben Milligan, 1979 (S. 344) und Steinhausen/Langer, 1977 (S. 77). Referenzen zu den einzelnen Verfahren gibt Cormack, 1971 (S. 331). - Für das ‚Single Linkage‘ ergibt sich tatsächlich  $d(G_r, G_p \cup G_q) = \min(d_{rp}, d_{rq})$ , da man für beliebige reelle Zahlen  $x, y$  schreiben kann:  $\min(x, y) = (x+y - |x-y|)/2$ .)

Eine an der Rekursionsformel orientierte Übersicht der agglomerativen Verfahren geben Eckes/Roßbach (1980, S. 67f.), sowie Steinhausen/Langer (1977, S. 76ff.). (Für die von Lance/Williams 1967 aufgrund der Rekursionsformel entwickelte ‚Flexible Strategie‘ schlagen die Autoren selbst einen Wert von  $\beta = -.25$  vor. Steinhausen/Langer (1977, S. 77) empfehlen das Intervall  $-.2 \leq \beta \leq -.4$ .)

Milligan (1979) verwendet die Rekursionsformel, um für bestimmte Parameterwerte  $\alpha_p$ ,  $\alpha_q$ ,  $\beta$  und  $\gamma$  nachzuweisen, daß die zugehörigen Verfahren die Eigenschaft haben, monotone (ultrametrische) Hierarchien (ohne Inversionen) zu liefern. Danach ist diese Eigenschaft gegeben, wenn  $\alpha_p + \alpha_q + \beta \geq 1$  mit  $\alpha_p, \alpha_q, \gamma \geq 0$  oder wenn  $\gamma < 0$  mit  $|\gamma| \leq \alpha_p, \alpha_q$  gilt.

Die Beachtung der Verfahrenseigenschaft, invariante Ergebnisse (bezüglich der Folge der Zerlegungen) bei monotoner Transformation der Proximitäten zu liefern (Johnson, 1967; Hubert, 1973), hat ihre Wurzeln in der nicht-metrischen mehrdimensionalen Skalierung (siehe z.B. Ahrens, 1974, S. 163ff.). Die Methoden des ‚Single-Linkage‘ und des ‚Complete-Linkage‘ erfüllen dieses Kriterium, da sie bei der Ermittlung der Distanzen zwischen einer ‚neuen‘, durch Fusion entstandenen Gruppierung  $G_p \cup G_q$  und den jeweils ‚anderen‘ Gruppierungen  $G_r$  lediglich ‚Kleiner-‘ bzw. ‚Größer-Abfragen‘ vornehmen, während durch alle anderen Verfahren ‚echte‘ arithmetische Rechenoperationen durchgeführt werden.

Fraglich ist allerdings, ob man dem Kriterium der ‚monotonen Invarianz‘ - wie bisher - eine besondere Dignität zuschreiben sollte. Es besagt zwar, daß die Verfahren zur Analyse nur die in den Daten vorhandene Ranginformation nutzen; dies bedeutet aber nicht, daß die Ergebnisse hinsichtlich der Repräsentation der (empirischen) Ranginformation auch optimal sind. Im Gegenteil, da ‚Single-Linkage‘ und ‚Complete-Linkage‘ nur die Information der jeweiligen

Extremwerte in den nächsten Analyseschritt transferieren, dürften die metrisch arbeitenden Verfahren in der Regel, auch bezogen auf nicht-metrische Kriteriumsmaße, wie z.B. Kruskals ‚Stress‘ oder Guttmans Maß(e) der schwachen Monotonie, Ultrametriken generieren, die die Ranginformation in den Proximitydaten besser abbilden. (Nebenbei sei auf die einfache Möglichkeit hingewiesen, allen HCA-Verfahren die Eigenschaft der Invarianz bei monotoner Transformation dadurch zuzueignen, daß man vor Verfahrensanwendung die Proximities rangiert, um dann die Ränge metrisch zu behandeln. Ein solches Vorgehen dürfte aber wohl als krude zu bezeichnen sein.)

Charakteristisch für die Methoden des ‚Single-Linkage‘ und des ‚Complete-Linkage‘ sind ihre **Verzerrungseigenschaften** („Spate-Distortion“, Lance/Williams, 1967; Hubert/Schultz, 1975). Das ‚Single-Linkage‘ neigt dazu, Objekte bzw. Gruppierungen sukzessive aneinanderzureihen (Kontraktion, ‚chaining-effect‘). Es tendiert zur Ausbildung einzelner großer Gruppen mit Durchmessern, die größer sein können als die Distanzen zwischen Objekten aus verschiedenen Gruppierungen. Das ‚Complete-Linkage‘ neigt zur Ausbildung gleichgroßer Gruppen; das zugehörige Dendrogramm macht zumeist einen ‚stark strukturierten Eindruck‘; damit gibt es Separationen von Objekten bzw. Gruppierungen vor, die relativ eng zusammenliegen können (Dilatation). - Diese Hinweise auf mögliche Artefaktquellen in den Verfahren gehen auf Lance/Williams (1967) zurück; sie wurden z.B. von Cormack (1971, S. 331) als zuwenig formalisiert kritisiert, Baker/Hubert (1975) aber konnten die Angemessenheit der Verfahrenscharakterisierung stützen, indem sie in einer Simulationsstudie die durch die Methoden produzierten Typen von Partitionen (Verteilungen der Gruppengrößen) untersuchten (siehe auch Hubert/Schultz, 1975).

Zur Beurteilung der **Leistung** der **Rekursionsformel** können die folgenden Gesichtspunkte herangezogen werden:

- Sie leistet eine exakte und kompakte Systematisierung der (bis 1967) bekanntesten agglomerativen HCA-Verfahren und entwickelt daraus eine ganze Klasse neuer heuristischer Methoden („Flexible Strategie“). Dieses beinhaltet allerdings auch die Gefahr der Einschränkung des Blickwinkels aufgrund des Eindrucks, alle denkbaren Methoden müßten durch die Parameter der Rekursionsformel beschreibbar sein.
- Sie leistet eine ökonomische prozeßbezogene Beschreibung der Analysen, indem sie verdeutlicht, daß während der Durchführung der Verfahren nicht ständig auf die ursprüngliche Matrix der Proximities  $d(o_k, o_l)$  zurückgegriffen werden muß; ausreichend ist die Untersuchung der Proximities zwischen den Gruppen  $d(G_p, G_q)$ , deren Anzahl geringer ist (Informationsreduktion). Dieser Gesichtspunkt war früher bei der Realisierung der Algorithmen auf elektronischen Rechnern von entscheidender Bedeutung.

schen Rechenanlagen von Belang (Speicherplatzbedarf), seine Beachtung ist aber jetzt, außer bei Problemen mit sehr großen Anzahlen von Objekten (z.B.  $n > 500$ ), nicht mehr zeitgemäß.

Es sollen nun **weitere agglomerative HCA-Verfahren** angesprochen werden, die nicht mehr mit Hilfe der Rekursionsformel charakterisierbar sind; sie verlangen während des gesamten Analyseprozesses die Untersuchung der ursprünglichen empirischen Proximitäten  $d(o_k, o_l)$ . Die hier benannten Methoden haben die Eigenschaft, daß ihre Ergebnisse bei monotoner Transformation der Proximitäten invariant bleiben; dabei rechnen sie nicht nur in den (jeweiligen) Extremwerten, wodurch die unangenehmen Verzerrungseigenschaften des ‚Single-‘ und ‚Complete-Linkage‘ teilweise vermieden werden können:

- (a) D'Andrade (1978) schlägt ein Verfahren vor, das auf den jeweiligen Stufen die Gruppierungen zusammenlegt, die nach dem, aus der nicht-parametrischen Statistik entlehnten, Kriterium der Mann-Whitney-U-Statistik maximal ähnlich bzw. minimal distant sind. Die Methode verwendet zur Fusionsentscheidung lediglich Auszählungen der Ergebnisse von Größenvergleichen für Proximitäten von Objektpaaren aus verschiedenen Gruppen (sie führt deshalb zu Dendrogrammen, die nicht numerisch evaluiert sind, sog. ‚bare trees‘ (Boorman/Olivier, 1973)).
- (b) Durch Setzung von ‚Funktion‘  $\equiv$  : ‚Median‘ in (F.I) entsteht ein Verfahren, das die Gruppen nach dem Kriterium des minimalen Medians der Distanzen von Objekten aus zwei verschiedenen Gruppen zusammenfaßt. (Diese Methode darf nicht mit der geometrisch orientierten ‚Median Method‘ (Nr. 6 der Übersicht) verwechselt werden.) (Weitere Verfahren ergeben sich durch Einsetzen anderer, insbesondere robuster Maße der zentralen Tendenz der Proximitäten von Objekten aus (zwei) verschiedenen Gruppen.) Speziell die Verwendung des Medians wurde, wie Johnson (1967) mitteilt, von J. D. Carroll (mündlich) vorgeschlagen.
- (c) Ergänzend ist an dieser Stelle auf ein Verfahren von Hubert (1973, S. 55), die ‚Objective Function Method‘ und auf die Klasse der ‚r-Diameter-Hierarchical-Clusterings‘ hinzuweisen (Hubert/Schultz, 1975; Hubert/Baker, 1977). Diese Methoden sind zwar ‚monoton invariant‘, aber nicht ‚konservativ‘, d.h. sie zeigen in unterschiedlichem Maße Verzerrungen.

Die unter (a) bis (c) angesprochenen heuristischen agglomerativen HCA-Verfahren zeigen zwar gegenüber ‚Single-‘ und ‚Complete-Linkage‘, wie von D'Andrade (1978) und Hubert/Baker (1977) für ihre Methoden ausgewiesen wird, die besseren Goodness-of-Fit Leistungen hinsichtlich des nicht-metrischen Goodman-Kruskal  $y$ , es kann aber (noch) nicht allgemein garantiert werden, daß die von den Verfahren generierten Resultate monoton hierarchisch, d.h. ultrametrisch sind.

### **3.3.3 Subdivisive Verfahren**

Für die Klasse der ‚absteigenden‘ HCA-Methoden, die von der gröbsten Partition  $Z(0) = 0$  ausgehen und die Zerlegungen sukzessive verfeinern, kann generell ausgesagt werden, daß konzeptuell alle Ansätze der Mengenzerlegung (Abschnitt 3.2) und der agglomerativen Verfahren (s.o.) als Heuristiken zur Konstruktion divisiver Methoden herangezogen werden können. Für zusammenfassende Darstellungen, wie der hier vorliegenden ist es daher verständlich, wenn der Diskussion zerlegender hierarchischer Methoden weniger Raum gegeben wird. Die geringe Verbreitung in der Anwendung ist darüber hinaus leicht aus den gewöhnlich immensen Anzahlproblemen der (binären) Partitionierung erklärt. Solange aber kriteriumsgeleitete Optimierungsverfahren zur Approximation von Ultrametriken an empirische Proximitymatrizen nicht allgemein verfügbar sind, sollte insbesondere den heuristischen Verfahren der sukzessiven Zerlegung erhöhte Beachtung zukommen. Dieser Empfehlung soll mit der folgenden Argumentation (nochmals) Nachdruck verliehen werden: Während agglomerative HCA-Verfahren zunächst in den kleinen Distanzen arbeiten und (Vor-)Entscheidungen treffen, gehen divisive Methoden anfangs von der Untersuchung der großen Distanzen bzw. geringen Ähnlichkeiten aus; es sind aber gerade die großen Entfernung, welche für die Global- oder Gesamtstruktur einer Objektmenge besonders prägend sind. Die Geltung dieser Aussage wird durch eine Untersuchung von Graef/Spence (1976, zitiert bei Kruskal, 1977) gestützt. Die Autoren konnten durch die Anwendung von MDS-Algorithmen nachweisen, daß die Streichung kleiner Distanzen aus der Proximitymatrix auf die Struktur von Punkten im Raum erwartungsgemäß weit geringer deformierend wirkt, als die Streichung großer Distanzen. Deshalb sollten Investitionen in die Analyse der größeren Unähnlichkeiten lohnend sein, zumal die zunehmende Verbreitung schneller Computer solchen intensiven Bemühungen entgegenkommt.

Im Einzelnen sei ergänzend auf folgende Ansätze der sukzessiven Zerlegung hingewiesen:

- (a) Hubert (1973) stellt (neben einem weiteren Verfahren) subdivisive Pendants zum agglomerativen ‚Single-‘ und ‚Complete-Linkage‘ zur Verfügung. Diese Methoden liefern ebenfalls invariante Resultate bei monotoner Transformation der Proximitymatrizen und sind zudem ökonomisch durchführbar.
- (b) Ein zum agglomerativen ‚Minimum Variante‘ Verfahren nach Ward (1963) (Methode 7 der Übersicht in 3.3.2) gehörendes subdivisives Pendant haben Edwards/Cavalli-Sforza (1965) vorgeschlagen. Ihm wird wegen der benötigten extremen Rechenzeiten z.B. von Eckes/Roßbach lediglich „illustrativer Wert“ (1980, S. 80) zugeeignet. Diese Einschätzung muß aber wohl revidiert werden, nachdem Herden/Steinhausen (1979 mit Fortran-Pro-

gramm) einen effektiven Branch-and-Bound Algorithmus zur Berechnung des globalen Minimums des Varianzkriteriums angeben.

- (c) Für die (in Abschnitt 3.3.2 angesprochene) Festlegung ‚Funktion‘=: ‚Median‘, bei der im subdivisiven Fall der Median der paarweisen Distanzen von (zwei) Gruppen für die Aufteilung maximiert wird, können - wie auch für andere komplexe Kriteriumsfunktionen - Austauschalgorithmen verwendet werden. (Diesem Vorgehen folgt Späth (1976) bei Vorliegen von Datenmatrizen  $[x_{ij}]$  für die Minimierung des  $L_1$ -Kriteriums: Summe der absoluten Abweichungen vom Median. Siehe auch Späth (1977) und Banfield/Basill (1977), sowie Späth/Müller (1979) für weitere Anwendungen (jeweils mit Fortran-Programmen).)
- (d) Eine interessante Beobachtung führte McQuitty (z.B. 1968) zur Entwicklung seiner ‚Iterative Intercolumnar Correlational Analysis‘: Korreliert man paarweise die Zeilen einer Korrelationsmatrix  $K_1$  zur Herstellung einer Korrelationsmatrix  $K_2$  und iteriert dieses Vorgehen, so tritt gewöhnlich („usually“, S. 466) ein Konvergenzphänomen ein:  $K_N = K_{N+1}$ . Die Matrix  $K_N$  hat dann nur noch Eintragungen von +1 und -1 und ist einer Zerlegung der Menge der Variablen eineindeutig zugeordnet. Wendet man dieses Vorgehen sukzessive auf die entstandenen Untergruppen bzw. deren zugehörige Untermatrizen an, so entsteht im (zumeist) erfolgreichen Fall eine Folge von (binären) Zerlegungen. Das Verfahren kann somit als heuristische subdivisive HCA angesehen werden. Es hat zumindest folgende Schwächen :
  - (1) Auf jeder Stufe wird die Zerlegung ohne jedes explizite Optimierungskriterium vorgenommen. Diese Eigenschaft teilt die Methode mit anderen Zerlegungsprozeduren.
  - (2) Es ist fraglich, ob die Produkt-Moment-Korrelation wegen ihrer Invarianz bei linearen Transformationen als Maß für die Übereinstimmung von Zeilen von Korrelationsmatrizen geeignet ist.
  - (3) Die Konvergenzeigenschaft des Vorgehens konnte, trotz der Begründungsbemühung von McQuitty/Clark (1968) bisher nicht bewiesen werden.
 Dazu sei hier mitgeteilt, daß (nach eigenen Untersuchungen) auch bei Verwendung verschiedener Distanzmaße zur iterativen Analyse dieser Art, nach einigen Schritten (mit Normierung) Konvergenz in dem Sinn eintritt, daß die resultierende Distanzmatrix eine Verteilung hoher gegenüber sehr niedrigen Eintragungen enthält, die einer Zerlegung zugeordnet ist. (Er-gänzend sei auf eine vergleichbare Technik von Bonner (1964) verwiesen, der Ähnlichkeitsmaße verwendete.) - Solche Proximitymaße von Proximities und ihre Leistungen zur Generierung disjunkter Hierarchien sollten verstärkt formal und im Rahmen von (vergleichenden) Evaluationsstudien untersucht werden. Die inneren Zusammenhänge dieser Vorgehensweise und ihre Effizienz für die Anwendung bedürfen der Klärung.

### **3.3.4 Evaluation, Anwendung und Weiterentwicklungen**

Evaluationsstudien zu HCA-Verfahren liegen m. W. ausschließlich für deren agglomerative Varianten vor. Zwei Typen von Studien lassen sich grob unterscheiden :

- (a) Tests nach dem **Modell der Mischverteilungen** gehen von einer Gesamtpopulation aus, die sich aus mehreren Subpopulationen mit unterschiedlichen Verteilungstypen, Mittelwert(svektor)en und/oder Kovarianzmatrizen zusammensetzt. Aus der Gesamtpopulation werden Stichproben generiert und den HCA-Verfahren unterworfen, zur Leistungsbewertung wird abschließend die Herkunft der Objekte (Punkte) aus den jeweiligen Subpopulationen mit ihrer Clusterzuordnung durch die Methoden über Kappa- oder Rand-Statistiken verglichen (Blashfield, 1976, Edelbrock, 1979, Edelbrock/McLaughlin, 1980, McIntyre/Blashfield, Milligan, 1980).

Bei insgesamt guten Zuordnungsleistungen sind die Verfahren (im Sinn einer Interaktion) differentiell sensitiv für verschiedene, durch das Populationsmodell spezifizierte Ausgangssituationen. Es werden also keine generellen Empfehlungen ausgesprochen. (Nach meinem Eindruck ergibt eine Übersicht leichte Vorteile für die ‚Average‘-Methoden und die ‚Minimum-Variante‘-Technik.) Relativ einhellig dagegen sind die Warnungen vor der Betrachtung nur einer Lösung, da die Verfahren in der Regel sehr verschiedene Resultate erbringen, die es jeweils genau zu untersuchen gilt.

(Für diesen Typ von Evaluationsstudien lassen sich die Leistungen der HCA-Verfahren mit denen der Mengenzerlegungsmethoden vergleichen. Dabei sind die Ergebnisse der für diesen Zweck speziell konstruierten Partitionierungsalgorithmen (z.B. aus der K-Means-Gruppe, Milligan, 1980) erwartungsgemäß besser, wenn ihnen nur plausible Startpartitionen (z.B. HCA-Verfahren als Heuristik) zur Verfügung gestellt werden.)

- (b) Untersuchungen zur Anfälligkeit von HCA-Verfahren bei **Vorliegen von Fehlern in den Daten** gehen von der Konstruktion ‚wahrer‘ metrischer oder ultrametrischer Proximitymatrizen aus. Ultrametriken werden dann mit aus verschiedenen Verteilungen stammenden zufälligen Fehlern überlagert und den HCA-Verfahren unterworfen. Zur Leistungsbewertung wird abschließend die Ranginformation in der ursprünglichen Metrik oder Ultrametrik mit der Ranginformation in den durch die Verfahren generierten Ultrametriken verglichen; dazu verwendet man Kendalls Tau oder Goodman-Kruskals Gamma (Baker, 1974; Cunningham/Ogilvie, 1972; D’Andrade, 1978; Hubert/Baker, 1977). Auch in diesen Untersuchungen ergibt sich eine Interaktion von Ausgangssituation und Methoden: Sind die ‚wahren‘ (Ultra-)Metriken ‚verkettet‘, so wird dies vom ‚Single-Linkage‘ besser reproduziert; ansonsten stimmen die Ergebnisse des ‚Complete-Linkage‘, vor allem aber die ‚mittleren‘ Verfahren, z.B. ‚Group-Average‘

und ‚Minimum Variante‘, deutlich besser mit den ursprünglichen Daten überein.

Da dem Anwender in der Regel nicht bekannt ist, inwieweit ‚Verkettungen‘ in seinen Proximitydaten vorliegen, bleibt ihm (wiederum) nur die Durchführung mehrerer HCA-Verfahren und die genaue Untersuchung ihrer Ergebnisse.

Evaluationsstudien führen also nicht zu generellen Empfehlungen. Vielmehr scheinen die heuristischen HCA-Verfahren für verschiedene Aspekte oder Merkmale der Daten sensibel zu sein, bzw. scheinen durch ihre Darstellungsresultate unterschiedliche ‚Aspekte‘ oder Merkmale der Daten hervorzuheben oder ihnen aufzuprägen.

Die für die ökonomische Anwendung von HCA-Verfahren notwendigen *Programme* wurden überwiegend für die agglomerativen Varianten publiziert, z.B. Anderberg (1973, S. 276ff.), Hartigan (1975, S. 215ff.), Späth (1975, S. 165ff.), Steinhausen/Langer (1977, S. 83ff.) Oder in Programmsysteme implementiert, z.B. BMDP (siehe Dixon, 1975, S. 307ff.) und Clustan (siehe Wishart, 1978, S. 31 ff.). Veröffentlichungen von Programmen zu den hierarchisch subdivisiven Verfahren finden sich nur vereinzelt, z.B. Späth (1975, S. 148 ff.), so daß man (vorerst) ersatzweise auf die unökonomische wiederholte Anwendung von Mengenzerlegungsprozeduren zurückgreifen muß (Hinweise zu Programmen, siehe Abschnitt 3.2). - Eine Diskussion und partielle Evaluation zu Computerprogrammen für hierarchische Clusteranalysen findet man bei Aldenderfer/Blashfield (1978). Ergänzend sei hier auf einen beachtenswerten Gesichtspunkt der Programmanwendung hingewiesen (der auch in der methodologischen Literatur zumeist übergangen wird): Generell können während der Durchführung von (heuristischen) Verfahren Eindeutigkeitsprobleme auftreten. Bei agglomerativen Verfahren z.B. können in einer Matrix von Unähnlichkeitsmaßen zwischen Gruppen zwei (oder mehrere) minimale Eintragungen vorliegen, mit  $d(G_p, G_q) = d(G_p, G_r)$ . Es ist nun mit Hilfe eines begründeten Entscheidungskriteriums die Frage zu beantworten, ob die Gruppe p mit der Gruppe q oder mit der Gruppe r fusioniert werden soll. Wird dieses Problem in einem Programm nicht explizit thematisiert, so nimmt dieses bei der Ausführung zumeist automatisch eine Entscheidung vor, die vom Benutzer unbemerkt bleibt. Das Verfahren erfüllt dann ein relevantes methodisches Kriterium nicht, nämlich das der Unabhängigkeit des Resultats von der Reihenfolge der Objekte (zu diesem und weiteren Kriterien siehe Wright, 1974). Ein positives Gegenbeispiel aber geben z.B. Steinhausen/Langer (1977, S. 83), die einen programmierten Fehlerindikator einführen.

In der Forschungspraxis stellen sich im Zusammenhang mit der Anwendung von hierarchischen Clusteranalysen in verschiedenen Phasen des **Vorgehens** eine Reihe von Problemen, deren Darstellung im folgenden mit entsprechenden Lösungsansätzen versehen wird:

Für vorliegende Daten stellt sich vor der Anwendung von HCA-Verfahren die Frage nach deren Angemessenheit für die Repräsentation der Proximitäten. Damit wird für eine **a-priori-Evaluation** nach einem Entscheidungskriterium gesucht, welches die ‚clusteriness‘ (siehe Sneath 1969, S. 263) der Daten beurteilt. - Für den Fall gegebener Datenmatrizen (Beobachtungswiederholungen x (numerische) Merkmale) haben Oldenbürger/Becker (1976, besser verfügbar in Oldenbürger/Schwibbe, 1980) eine Strategie entwickelt, die eine statistische Prüfung der Clusterbarkeitshypothese gestattet. (Zu einem ähnlichen Vorgehen führt ein unabhängiger Vorschlag Sodeurs (1976) im Anschluß an Vogel (1975).) Dazu wurde eine Klasse von Maßen definiert, die die Abweichung der Proximitäten, z.B. Distanzen, von der ultrametrischen Ungleichung quantifizieren. (Passager wird ein einfacheres Maß auch z.B. von Carroll/Pruzansky 1980, S. 113 erwähnt.) Die Konstruktion des statistischen Tests erfolgte unter Verwendung der ‚Philosophie‘ der Randomization-Tests (Edgington 1969, 1980); dazu wurde eine geeignete Randomisierungsbasis gewählt, in der die Verteilungen der Variablen und damit ihre Kenngrößen erhalten bleiben. Für die statistische Entscheidung wird dann die Abweichung von der ultrametrischen Ungleichung für die empirischen Distanzen (hinsichtlich einer gewählten Modalität) mit der Verteilung der Abweichungen von der ultrametrischen Ungleichung für Distanzen aus zufälligen Datenmatrizen verglichen. (Diese werden durch zufällige Permutation der ursprünglichen Daten innerhalb der Variablen über die Beobachtungswiederholungen generiert.) - Für den Fall direkt erhobener Proximitymatrizen schlagen Sattath/Tversky (1977) die Untersuchung der Verteilung der Proximitäten vor. Sie konnten nämlich zeigen (Skewness-Theorem, 1977, S. 342ff.), daß die Verteilungen von Distanzen zwischen Punkten, die in euklidischen Räumen liegen, eher rechtsschief bzw. linkssteil sind, während Verteilungen von Distanzen zwischen Punkten in additiven Bäumen, z.B. Dendrogrammen, linksschief bzw. rechtssteil sind. (Dieser Ansatz ist allerdings bisher nicht mit einem statistischen Prüfverfahren gekoppelt, was aber durch Konstruktion eines Randomization-Tests für bestimmte Fälle relativ leicht geleistet werden kann. Dabei sind dann die tripelweisen Einschränkungen der Distanzen über die Dreiecksungleichung zu berücksichtigen.)

Hat man sich für die Durchführung der hierarchischen Clusteranalyse entschieden, so gibt es a-priori keine spezielle Methode der Wahl, deren Resultat eine optimale Repräsentation der Proximitäten garantiert. Es sollten also vielfältige, möglichst auch subdivisive Verfahren eingesetzt werden. Nach deren Anwendung stellt sich dann die Frage nach einer quantitativen **a-posteriori-Evaluation** der Resultate und damit einer nachvollziehbaren Auswahl der besten Repräsentation(en). Es gilt also, geeignete Goodness-of-Fit Kriterien für hierarchische Gruppierungen zu definieren (siehe z.B. Gower, 1975; eine Liste gibt Cormack, 1971, S. 338):

- Soll den Proximitydaten nur der Charakter von **Ranginformationen** unterstellt werden, so bietet sich als globales Maß für die Güte der Anpassung der ursprünglichen Proximities durch die aus den Dendrogramm rekonstruierten ultrametrischen Ähnlichkeits- bzw. Unähnlichkeitsmaße Goodman-Kruskals  $\gamma$  an (ein Rechenbeispiel gibt Lienert, 1978, S. 554ff.). Selbstverständlich dürfen auf diesen Index die üblichen statistischen Prüfverfahren nicht mehr angewendet werden. (Für kleine  $n(<16)$  und die speziellen Verfahren des agglomerativen ‚Single-‘ und ‚Complete-Linkage‘ gibt Hubert (1974) Tabellen, die eine zufallskritische Beurteilung von  $\gamma$  gestatten. Wegen der Verfahrensabhängigkeit dieses Vorgehens ist die statistische a-priori-Prüfung der globalen Clusterbarkeitshypothese durch (approximative) Randomization-Tests wohl vorzuziehen.)
- Sind die Proximitymaße **metrisch** konstruiert worden oder wird ihnen metrischer Charakter zugeeignet, so bedarf es eines Zwischenschrittes vor der Anwendung globaler Goodness-of-Fit Kriterien. Wählt man z.B. die (kopfnetische) Korrelation  $r_{(empd_{kl}, ultd_{kl})}$  (Hubert/Baker, 1977b), das ist die Produkt-Moment-Korrelation zwischen den Eintragungen der Dreiecksmatrizen (ohne Diagonale), oder die  $L_p$ -Abstandsnorm  $\sum_{k<l} |empd_{kl} - ultd_{kl}|^p$ , mit  $p=2$  als globales Maß für die Abbildungstreue der ursprünglichen Daten durch die konstruierte Ultrametrik, so ist dieser Index nur bei zgl. des ‚Group-Average‘ oder seines divisiven Pendants als fair anzusehen (Farris, 1969). Alle anderen Verfahren generieren nämlich global und/oder partiell gedehnte oder gestauchte Dendrogramme bzw. Ultrametriken, so daß ein Vergleich der einzelnen Strukturierungsvorschläge (Dendrogramme) erst nach Beseitigung dieser Verfahrenseinflüsse durch Berechnung einer optimal angepaßten Ultrametrik  $ultd_{kl}$  erfolgen sollte. Dazu stehen die Techniken der numerischen Approximation gegebener Baumstrukturen an Proximitymatrizen zur Verfügung, wie sie etwa bei Eisler (1973) oder Roskam (1975, S. 49f.) für die generelle Anwendung beschrieben werden. Im hier zur Debatte stehenden Fall aber reicht es aus, die ( $L_2$ -)Knotenhöhe jeweils durch Mittelung der empirischen Proximities anzugeben, deren zugehörige Objektpaare durch den Knoten erst zusammengeführt werden. (Bei Wahl der robusteren  $L_p$ -Abstandsnorm mit  $p = 1$  wäre hier die Angabe des Medians angemessen.) Nach dieser numerischen Anpassung können die strukturellen Verfahrensvorschläge (Dendrogramme) verbessert graphisch dargestellt und ihre Abbildungsleistung durch Berechnung von z.B.  $r_{(empd_{kl}, ultd'_{kl})}$  quantitativ bewertet werden.

Hat man nach Berechnung der Globalmaße eine begründete Auswahl unter den Repräsentationen getroffen, so verbleibt als Aufgabe die spezifischere Untersuchung der Abbildungsleistung des HCA-Resultats. Diese kann durch Inspektion der Rangplatzdifferenzen oder der Residuen  $empd_{kl} - ultd'_{kl}$  erfolgen; das Vorzeichen solcher Werte gibt dann an, ob die jeweiligen Objektpaare

durch die clusteranalytische Darstellung als zu stark separiert oder als zu eng zusammengehörig ausgewiesen werden; die absolute Größe solcher Abweichungswerte zeigt auf, an welchen Stellen die kompakte Repräsentation durch hierarchische Gruppierung erhebliche ‚Fehlleistungen‘ zeitigt. - Eine solche (exploratorische) Bemühung kann einerseits undifferenzierte Globalinterpretationen von Dendrogrammen vermeiden helfen, andererseits könnte die Verteilung (das ‚Pattern‘) der Abweichungen Hinweise auf (weitere) alternative Repräsentationsmodi der Proximitydaten liefern.

Sollen in der Anwendung verschiedene hierarchische Strukturierungen verglichen werden, so ist dazu auf folgende Ansätze hinzuweisen:

- Boorman/Olivier (1973) definieren eine Reihe von Metriken über Folgen von Partitionen bzw. über Ultrametriken, die z.B. zur Auswahl eines ‚mittleren‘, maximal konsistenten Dendrogramms herangezogen werden sollen.
- Zur konfirmatorischen statistischen Prüfung der Übereinstimmung a-priori formulierter hierarchischer Strukturhypothesen mit den Resultaten der Anwendung von HCA-Verfahren auf (unabhängige) Daten setzen Hubert/Baker (1977) das äußerst flexible Paradigma des ‚quadratischen Assignierens‘ ein (Lawler, 1975; Hubert/Schultz, 1976). Diese generelle Strategie der Datenanalyse ist über die Untersuchung von Permutationsverteilungen eng mit der Methodologie der Randomization-Tests verknüpft. Sie gestattet auch eine konfirmatorisch zufallskritische Beurteilung der Übereinstimmung von (z.B. nach Verfahrensanwendung ultrametrischen) Proximitymatrizen aus verschiedenen Untersuchungsgruppen (Subsamples). Gleichzeitig liefert die Theorie des quadratischen Assignierens eine Grundlage zur probabilistischen Evaluation der von Boorman/Olivier (1973) vorgeschlagenen Distanzmaße für Folgen von Partitionen. - Bei der Anwendung dieser komfortablen Evaluations- und Prüfstrategie ist zur Vermeidung von Artefaktinterpretationen grundsätzlich zu beachten, daß die zu vergleichenden Proximitymatrizen unabhängiger Herkunft sind. (So wäre z.B. die ‚Prüfung‘ der kophenetischen Korrelation zwischen empirischer Proximitymatrix und der nach Verfahrensanwendung rekonstruierten Ultrametrik unzulässig.) - Erhebliche Weiterentwicklungen des Paradigmas des quadratischen Assignierens durch die Einführung und Untersuchung des Vergleichs genereller Proximityfunktionen werden von Hubert/Baker (1977), Hubert (1978) und Hubert/Baker (1979) angeregt. Im hier gegebenen Diskussionszusammenhang ist die Möglichkeit des simultanen Vergleichs K unabhängiger Proximitymatrizen von besonderem Interesse (Hubert, 1979c; siehe auch Hubert, 1979a, 1979b). An dieser Stelle muß aber ebenso auf die Ansätze zur Prüfung von Strukturhypothesen zu Korrelations- und Kovarianzmatrizen in den Arbeiten Jöreskogs (z.B. 1978) hingewiesen werden (siehe auch Steiger, 1980).

Abschließend sei hervorgehoben, daß die Ausführungen in diesem Kapitel sich vornehmlich auf die Anwendung von HCA-Verfahren zur Konstruktion von Folgen von Partitionen mit Gruppen lagehomogener Objekte bezogen. Diese Einschränkung aber ist weder notwendig noch sinnvoll. Nach Definition geeigneter Proximitymaße und/oder Optimierungskriterien können Verfahren, die hierarchische Repräsentationen generieren, für **verschiedenste Zielsetzungen** verwendet werden. Als Beispiele für weitere Entwicklungen seien die folgenden Ansätze genannt:

- Bei Vorliegen nominaler Datenmatrizen kann die hierarchische Entropieanalyse der ‚Canberra-Schule‘ eingesetzt werden (z.B. Williams/Lance, 1977).
- Wendet man die Hierarchisierung auf allgemeinere Problemstellungen der Mengenzerlegung an, wie sie zum Abschluß des Abschnitts 3.2 aufgewiesen werden (z.B. Aufteilung nach Distanz zu Hyperebenen, ‚Klassifikationsansatz‘), so erhält man weitere Verfahrensklassen, für die in erster Linie subdivisives Vorgehen adäquat sein wird.
- Die Aufhebung der Einschränkung, hierarchische Clusteranalysen auf symmetrische Proximitymatrizen anzuwenden (ohne diese einfach durch Mittelung korrespondierender Eintragungen zu symmetrisieren,) nimmt Hubert (1973) vor (siehe auch Hubert, 1976 und Baker/Hubert, 1977).
- In Abhebung von der sonst üblichen ‚ungerichteten‘ Clusteranalyse entwickelten Kleiter/Timmermann (1977) einen Algorithmus zur gerichteten hierarchischen Voraussetzungs-Struktur-Analyse. Eine ausführliche Darstellung dieser inhaltlich-theoretisch interessanten Repräsentationsverfahren am Beispiel der Analyse sachlogischer und empirischer Lernvoraussetzungen (Lernwege) geben Kleiter/Petermann (1977, mit Computerprogrammen, siehe auch Kleiter, 1980).

In den folgenden Abschnitten werden einige Ansätze dargestellt, die den Rahmen der Anwendung clusteranalytischer Verfahren auf die Untersuchung weiterer diskreter mathematischer Strukturen ausdehnen.

### 3.4 Baumrepräsentationen und hybride Modelle

Zusammenhängende, kreislose Graphen werden als Bäume bezeichnet (siehe z.B. Berge, 1976, S. 24f.; Harary, 1974, S. 42f.). Die Verbindung dieses Konzepts mit der Clusteranalyse ergibt sich insbesondere aus der Möglichkeit, die durch Anwendung von Verfahren der hierarchischen Clusteranalyse generierten Dendrogramme als spezielle Bäume anzusehen. Die daran anknüpfbaren Weiterentwicklungen der Baumrepräsentation von Proximitymatrizen (Patrinos/Hakimi, 1972) sollen im folgenden kurz angesprochen werden.

Faßt man Dendrogramme als ultrametrische Bäume auf (im Unterschied zur nicht-metrischen Interpretation als ‚bare trees‘, siehe Boorman/Olivier, 1973), so sind folgende Einschränkungen charakteristisch:

- (a) Alle Objekte erscheinen als Endpunkte bzw. terminale Knoten des Baums und
- (b) die Entfernung zwischen allen Objekten aus elementfremden Gruppierungen sind gleich groß.

Lockert man die Restriktionen und verlangt lediglich, daß den Kanten des Baumes positive Gewichtszahlen zugeordnet werden, deren Addition die Länge des Weges (Kantenfolge) zwischen den jeweiligen Objektpaaren angibt, so entstehen die sogenannten freien oder additiven Bäume. Die durch sie abgebildeten Distanzen folgen der sogenannten Vierpunktmetrik (Bunemann, 1971):

Für alle  $o_k, o_l, o_r, o_s \in O$  gilt:

$$(D. 6) \quad d(o_k, o_l) + d(o_r, o_s) \leq \max \{d(o_k, o_r) + d(o_l, o_s), d(o_k, o_s) + d(o_l, o_r)\}$$

(vgl. Abschnitt 2; es gilt ferner: (D. 5)  $\Rightarrow$  (D. 6)  $\Rightarrow$  (D. 4)). Algorithmen zur Repräsentation von Proximitymatrizen durch additive Bäume sowie Anwendungsbeispiele geben Cunningham (1974, 1978), Sattath/Tversky (1977), Carroll/Chang (1973) und Carroll (1967) sowie Carroll/Pruzansky (1980). In engstem Zusammenhang mit diesem strukturellen Ansatz stehen die fundamentalen theoretischen Diskussionen des Feature-Modells der Ähnlichkeit durch Tversky (1977) und des probabilistischen Choice-Modells der Präferenz durch Tversky/Sattath (1979). Colonius/Schulze (1977, 1979) geben notwendige und hinreichende Bedingungen für die Baumrealisierbarkeit einer empirischen, nicht-numerischen Struktur an und schlagen eine Konstruktionsmethode vor.

- Boyd/Wexler (1973) entwickeln eine Komponentenanalyse strukturierter Bäume (mit ungerichteten Kanten) und wenden sie auf die semantische Analyse von Verben an. - Läßt man für additive Bäume die Einschränkung fallen, daß die positiv gerichteten Kanten ungerichtet sind und ersetzt diese durch zwei gerichtete Kanten mit verschiedenen Gewichten, so erhält man sogenannte gewichtete bidirektionale Bäume. Cunningham (1978) benennt notwendige und hinreichende Bedingungen für deren Existenz und Eindeutigkeit und zeigt die Möglichkeiten der Repräsentation unsymmetrischer Proximitymatrizen durch bidirektionale Bäume an einem Beispiel aus dem Bereich des ‚sentence memory‘ auf.

Erste Vorschläge zur Repräsentation von Proximitydaten durch **gemischte bzw. hybride Modelle** gehen auf Degerman (1970, 1972) zurück. Diese kombinieren klassifikatorische Darstellungen mit denen der mehrdimensionalen Skalierung. (So unterscheiden sich beispielsweise Tiere einerseits hinsichtlich diskreter Charakteristika (z.B. ‚ist Säugetier‘ oder ‚ist nicht Säugetier‘), andererseits hinsichtlich kontinuierlicher Merkmale (z.B. Größe).) In der erweiterten

Fassung von Carroll (1976) werden gegebene Proximitymatrizen, z.B. von Unähnlichkeiten  $D_{\text{emp}}$ , als additive Kombinationen verschiedener (theoretischer) Proximitymatrizen dargestellt:

$$\text{HM: } D_{\text{emp}} \simeq D_1 + D_2 + D_3 + \dots + D_m$$

( $\simeq$  steht für ‚approximiert nach einem Optimalitätskriterium‘). Dabei stehen die  $D_1$  bis  $D_m$  für Proximitymatrizen, die z.B. Abstände aus (mehr)dimensionaler Skalierung, ultrametrischen Bäumen und/oder additiven Bäumen darstellen. Den überzeugenden Erfolgen dieses generellen Ansatzes bei der Repräsentation exemplarischer Daten (siehe auch Carroll/Pruzansky, 1980) und der Verfügbarkeit algorithmischer Lösungen, insbesondere durch alternierende Kleinst-Quadrat-Approximationen (siehe dazu auch die Referenzen in Young/de Leeuw/Takane, 1980), stehen z.Z. noch Probleme gegenüber, die auf die immense Flexibilität der Repräsentation und die entsprechend große Zahl der zu schätzenden Parameter zurückgehen; sie betreffen vor allem die Eindeutigkeit und Identifizierung der Modelle sowie deren Prüfbarkeit.

### 3.5 Überlappende Gruppierung

Die **bisher** in diesem Kapitel behandelten Ansätze zur Erzeugung homogener Gruppen verlangen, daß diese paarweise elementenfremd bzw. disjunkt sind:

Für alle  $p, q$  gilt:  $G_p \cap G_q = \emptyset$ .

Zwar sind in der Anwendung deutlich überwiegend Verfahren eingesetzt worden, die diese Einschränkung vornehmen, es lassen sich aber leicht Fälle angeben, die eine Einteilung von Objekten in einander ausschließende Klassen unangemessen erscheinen lassen. (So stellt z.B. die Liste wissenschaftlicher Publikationen unter sogenannten ‚key-words‘ im Index der ‚Psychological Abstracts‘ eine nichtdisjunkte Gruppierung dar; es wäre nicht praktisch, würde man jede Arbeit nur unter genau einem Schlagwort finden können (Zerlegung).) In der methodologischen Literatur findet sich eine Fülle von Verfahren, die für eine Objektmenge mit gegebener (symmetrischer) Proximitymatrix zumindest eine Menge  $A = \{G_1, G_2, \dots, G_p, G_q, \dots\}$  generieren, deren Elemente nichtdisjunkte Gruppierungen sind. Da diese Konstruktionsmethoden in der Regel ohne Optimalitätskriterium arbeiten, müssen sie als heuristisch bezeichnet werden. Neben dem notwendig hohen Aufwand liegt wohl auch in dieser Eigenschaft ein Grund, warum sie in der Anwendung bisher so wenig Beachtung fanden: Es liegt im allgemeinen kein Maßstab vor, nach dem alternative Klassifikationssysteme  $A_t$  als mehr oder weniger angemessen zur Repräsentation einer Objektmenge und deren Proximities beurteilt werden können. Von gegenwärtigen Verfahrensentwicklungen aber wird dieser Gesichtspunkt stärker berücksichtigt.

Die wohl ausführlichste Übersicht der theoretischen Grundlagen und der vielfältigen Verfahren zur überlappenden Gruppierung gibt Bock (1974, S. 316ff.). - In diesem Abschnitt wird dagegen die kurze Diskussion auf drei zentrale Ansätze eingeschränkt: die Untersuchung maximaler Cliques, die Generierung überlappender Cluster nach Jardine/Sibson und das ADCLUS-Modell von Shepard/Arabie.

Für die Betrachtung der Aufgabenstellung unter dem Stichwort ‚**maximale Cliques**‘ sei zu einer Objektmenge eine Matrix von Unähnlichkeiten  $[d_{kl}]$  vorgegeben. Unter einer Clique der Stufe c (cut) versteht man eine Menge  $G_p \subset \mathbf{0}$ , deren sämtliche Paare von Elementen einander nicht unähnlicher sind als durch c angegeben wurde:

Für alle  $o_k, o_l \in G_p$  gilt  $d_{kl} \leq c$ .

„Maximal“ heißt die Clique  $G_p$ , wenn es in der Restmenge  $\mathbf{0} \setminus G_p$  kein weiteres Objekt  $o_r$  gibt, dessen Unähnlichkeiten zu allen Objekten in der Clique kleiner/gleich c sind. Auf beinahe natürliche Weise führen diese Begriffsbestimmungen zu einem - auf den ersten Blick - einfachen heuristischen Konstruktionsverfahren überlappender Gruppierungen: (1) Man legt einen geeigneten Wert c fest und (2) sucht alle maximalen Cliques der Stufe c. Für beide Schritte gibt es algorithmische Lösungsansätze:

- (1) Faßt man beispielsweise die Unähnlichkeiten als metrische Größen auf, so kann ein optimaler ‚cut‘ z.B. mittels des Fisher-Algorithmus (siehe Abschnitt 3.2) erfolgen.
- (2) Für das nicht-triviale Problem der Feststellung aller maximalen Cliques der Stufe c steht dann z.B. der effektive ‚branch-and-bound‘-Algorithmus von Bron/Kerbosch (1973, weitere Angaben bei Shepard/Arabie, 1979, S. 99) zur Verfügung.

Allerdings ist dieses praktikable Vorgehen wohl als recht vordergründig anzusehen, da zu den genannten Schritten allgemeinere Probleme korrespondieren, die bei der Herstellung von Verfahren der nichtdisjunktiven Clusteranalyse zentral sind:

- (1) Wieviele Werte  $c_t$  sollen verwendet werden, um möglicherweise quasi-hierarchische Varianten des Verfahrens zu erzeugen?
- (2) Warum soll man sich auf den Gruppierungsbegriff der ‚maximalen Clique‘ festlegen? (Dieser garantiert zwar, daß der Durchmesser einer Gruppierung höchstens  $c_t$  ist (wie bei HCA-Verfahren das ‚Complete Linkage‘), man könnte aber ebenso begründet Gruppen generieren, deren Elemente zu allen anderen Objekten der Gruppe z.B. höchstens eine durchschnittliche Unähnlichkeit von  $c_t$  aufweisen.)

Bei Betrachtung der ersten Fragestellung ist folgende Argumentation relevant: Zu einer gegebenen Matrix von Unähnlichkeiten können sinnvoll höchstens  $\binom{n}{2}$  Cutwerte  $c_t$  untersucht werden. Geht man diese der Reihe nach durch und generiert die (z.B.) jeweils neu entstehenden maximalen Cliquen, so stellt dieses Mengensystem eine Repräsentation der Proximitymatrix ohne Informationsverdichtung dar, also eine reine Datendeskription. (Die Begründung der Gültigkeit dieser Aussage kann parallel zum Beweis des wichtigen Dekompositionstheorems der ‚Fuzzy Relations‘, siehe Kaufman (1975, S. 67ff. und s. 387ff.), erfolgen.) Eine strukturierte quasi-hierarchische Repräsentation von Proximities entsteht deshalb erst durch Einführung **spezieller Restriktionen**. Da es dafür verschiedene Möglichkeiten gibt, schlagen u.a. Jardine/Sibson (1968, 1971, S. 59ff., sieht auch Bock, 1974, S. 423ff.) drei Verfahrensklassen vor, deren bekannteste das sogenannte  $B_k$ -Clustering ist (siehe auch Cole/Wishart, 1970; Röhl, 1974, 1975; Day, 1977). Für diese Methoden ist als Einschränkung die erlaubte Anzahl von Elementen in den Überlappungen ( $k-1$ ) charakteristisch. Wegen ihrer engen Beziehung zu dem von Jardine und Sibson (1971) favorisierten hierarchischen ‚Single Linkage‘ Verfahren setzen sich diese Methoden allerdings der Kritik aus, Verkettungseffekten zu unterliegen. Listen von Fortran-Programmen geben Jardine/Sibson (1971, S. 227ff., u.a. auch zu Sibsons C-Verfahren, welche die erlaubte Größe des Durchmessers der Überlappung einschränken); die  $B_k$ -Methoden sind auch in Wisharts Programmpaket Clustan, 1978, S. 65ff., enthalten. Anwendungen dieser Verfahren sind bisher kaum bekannt (siehe aber Friendly, 1977).

Ein der zweiten Fragestellung zugeordneter Ansatz zur Generierung überlappender, nicht notwendig hierarchischer Gruppierungen stammt von Shepard/Arabie (1979). Das ‚additive clustering‘ geht von Ähnlichkeitsmaßen in der Proximitymatrix aus und repräsentiert diese durch ein einfaches Modell:

$$\hat{s}_{kl} = \sum_j w_j p_{kj} p_{lj}$$

Darin sind die  $w_j$  positive Gewichte, die die Sättigung einer Eigenschaft  $j$  angeben, während die  $p_{kj}$ ,  $p_{lj}$  zeigen, ob z.B. ein Objekt  $k$  die Eigenschaft  $j$  besitzt ( $p_{kj}=1$ ) oder nicht ( $p_{kj}=0$ ). Die Ähnlichkeiten zwischen den Objektpaaren werden also als Kombinationen der Gewichte diskreter Objektmerkmale dargestellt. Einen Algorithmus zur Lösung dieses Dekompositionsproblems unter Verwendung eines Maßes der erklärten Varianz als Optimierungskriterium entwickelten Shepard/Arabie (1979) unter dem Namen ADCLUS. Eine effizientere Alternative beschreiben Arabie/Carroll (1980). Sie verwenden Techniken der ‚mathematischen Programmierung‘ zur Lösung des Optimierungsproblems: MAPCLUS. Dabei scheinen sie im allgemeinen zu sparsameren Lösungen zu gelangen, d.h. zur Repräsentation der Ähnlichkeiten werden weniger Attribute benötigt. In den zitierten Arbeiten werden zahlreiche Beispiele für Ergebnisse der Verfahrensanwendung auf empirische Daten gegeben.

ben. Schließlich erweitern Arabie/Carroll (1980) das ADCLUS-Modell auf die Möglichkeit der Analyse individueller Differenzen (INDCLUS). - Geeignete statistische Prüfverfahren für diese Ansätze sind noch zu entwickeln. Die Residueninspektion im Anwendungsfall ist selbstverständlich.

### 3.6 Cluster in Datenmatrizen

Aufgabenstellung dieser Verfahren ist das ‚Auffinden‘ von Blöcken, Feldern, ‚Pattern‘ oder Strukturen gleicher oder ‚ähnlicher‘ Daten in  $n \times m$  - Matrizen  $[x_{ij}]$  (Beobachtungswiederholungen  $x$  Merkmale) durch geeignete Umordnung ihrer Zeilen und/oder Spalten. Während die in den bisherigen Abschnitten betrachteten Vorgehensweisen Gruppen ähnlicher Objekte aus **einer** Grundmenge konstruieren, untersuchen clusteranalytische Verfahren zur Strukturierung von Datenmatrizen Mengen von Objekten hinsichtlich **zweier** Modalitäten (simultan).

Kritisch für die Wahl bzw. Einsatzmöglichkeit solcher Verfahren ist die Beantwortung der Frage nach der Vergleichbarkeit kategorialer und/oder numerischer Merkmalsausprägungen (Hartigan, 1972, S. 124). Der genuine Anwendungsfall der Clusteranalyse von Datenmatrizen ist gegeben, wenn Vergleichbarkeit der Ausprägungen hinsichtlich der Merkmale **und** der Beobachtungswiederholungen vorliegt, bzw. durch geeignete Transformation (z.B. Standardisierung) hergestellt werden kann (Beispiele: (a) Matrix Bundesländer x Geburtsjahrgänge mit Eintragungen, die den prozentualen Anteil der Personen mit Abitur angeben, (b) Personen x Items eines Tests mit den Ausprägungen ‚gelöst‘ (= 1) und ‚nicht gelöst‘ (=0)).

Sind die Ausprägungen hinsichtlich einer Modalität (z.B. zwischen den Klassen verschiedener nominaler Merkmale) inkompatibel, so ist zumeist auf die in Abschnitt 3.1 bis 3.4 beschriebenen clusteranalytischen Verfahren zurückzugreifen (siehe ergänzend z.B. Hartigan, 1975, S. 257, aber auch Generalisierungsversuche ebd., S. 288ff. und deren Weiterentwicklung S. 299ff.).

Liegen bimodal vergleichbare Daten vor, so kann die Problemstellung der Verfahren der sogenannten direkten Clusteranalyse (Hartigan) oder der Konstruktion von Blockmodellen (Arabie/Boorman/Levitt, 1978; White, 1977) wie folgt formuliert werden: Gesucht ist sowohl eine Permutation der Zeilen- als auch der Spaltenindizes, so daß die (umgestellte) Datenmatrix bezüglich (der Einfachheit) ihrer inneren Strukturierung optimal ist. Beispiele für Zielfunktionen sind: (a) die Anzahl identischer Ausprägungen benachbarter Eintragungen bei kategorialen Merkmalen oder (b) die Summe der absoluten Differenzen benachbarter Eintragungen bei numerischen Merkmalen. (Zu beachten ist hier die (in der Literatur vernachlässigte) Möglichkeit der Einführung von ‚Fen-

stern' verschiedener Größe.) - Essentieller Bestandteil solcher Optimierungsverfahren ist die Konstruktion von Reihenfolgen (Seriation), deren algorithmische Untersuchung als Standardproblem des ‚Operations Research‘ wohlbekannt ist (Müller/Merbach, 1970; Christofides, 1975, 1975b, S. 236ff.) und erst durch Hubert (1976), Baker/Hubert (1977) und Hubert/Baker (1978) in die methodologische Diskussion der Psychologie eingebbracht wurde.

Beispiele für die Verfahrensanwendungen der direkten Clusteranalyse bzw. der Blockmodellierung für empirische Daten geben Hartigan (1975, S. 262 und S. 283 f.) und Arabie/Boorman/Levitt (1978).

Programmsprachliche Formulierungen von Verfahren (ohne besondere Hervorhebung des generellen Seriationsansatzes) teilt Hartigan (1975, S. 271 ff. und S. 294ff.) mit. Von diesem Autor stammt auch ein Programm, welches in das BMDP-Paket (Dixon, 1975, S. 339ff.) unter dem Namen ‚Block-Clustering‘ implementiert wurde. (Dabei handelt es sich wohl um eine ältere Version; es ist aber auf die instruktive Erklärung des Verfahrens (S. 352f.) und den ‚Sample-output‘ (S. 354f.) hinzuweisen.)

Die Verfahren der direkten Clusteranalyse von Datenmatrizen bzw. des ‚Block-Clustering‘ stellen insbesondere für exploratorische Zwecke hochinteressante Instrumente dar: sie strukturieren die ursprünglichen Daten und stützen die Anschauung. Da darin auch Gefahren liegen, ist die Wichtigkeit probabilistischer und statistischer Evaluationen besonders hervorzuheben (Ansätze geben z.B. Hartigan, 1975, S. 285 und White, 1977). Abweichungen der Daten von ‚Modellprädiktionen durch die Blöcke‘ sind in den Resultaten der Verfahrensanwendung (‚Print-outs‘) direkt ersichtlich.

#### *4. Diskussion und Ausblick*

Obwohl sich die vorliegende Arbeit in der Darstellung auf die Hauptgesichtspunkte clusteranalytischer Verfahren einzuschränken hatte, ist wohl auch hier deutlich geworden: Es gibt eine immense Vielfalt von Zielsetzungen, Problemstellungen, Ansätzen, Methoden und Verfahrensvarianten zur Clusteranalyse. - Dem steht in der Forschungspraxis z.Z. ein erheblich eingeschränkter Gebrauch von hauptsächlich ca. drei Verfahren der agglomerativen hierarchischen Clusteranalyse gegenüber. - Will man diese Diskrepanz erklären, so wäre der Verweis auf die generelle Angemessenheit der HCA-Verfahren für die Klärung Verschiedenster inhaltlich-theoretischer Fragestellungen sicher nicht akzeptabel. Daneben scheinen auch Argumente, die die ökonomische Durchführbarkeit dieser Methoden und die mangelnde Verfügbarkeit von Computerprogrammen zu alternativen Verfahren benennen, allein nicht überzeugen zu können.

Vermutlich kommen zumindest zwei Merkmale der Diskussionsentwicklung in der Literatur hinzu: (a) Nach dem Aufweis des methodischen Paradigmas durch den Artikel von Johnson (1967) sind die agglomerativen HCA-Verfahren als ‚Allroundinstrumente‘ (ähnlich dem ‚Little Jiffy‘ in der Faktorenanalyse, d.i. Hauptkomponentenextraktion und Varimaxrotation) üblich geworden. (b) Die ‚Landschaft‘ alternativer Ansätze und Verfahren in der methodologischen Literatur ist wegen ihres Umfangs und Aspektreichtums zu komplex und damit unüberschaubar geworden. Sie bietet kein geordnetes Bild, sondern macht eher einen diffusen Eindruck, der dem interessierten Anwender einen zielgerichteten Zugriff nicht erlaubt.

Ein Beleg für die Unübersichtlichkeit der mathematischen Darstellungen ist das ‚Sprachengewirr‘ der Formalisierungen. So findet man z.B. verbandstheoretisch (Leuschner, 1973), mengentheoretisch (Hubert, 1977a), matrixtheoretisch (Kim/Roush, 1978), graphentheoretisch (Hubert, 1974, 1977b; Matula, 1977) und schließlich ‚Fuzzy Set‘-theoretisch (Negoita/Ralescu, 1975, S. 169ff.; Yeh/Bang, 1975; Zadeh, 1977) geprägte Präsentationen der Clusteranalyse. - Ein weiteres Merkmal der methodologischen Literatur, welches die Orientierung erschwert, ist die vorwiegende **Betonung der rein verfahrenstechnischen Aspekte** des Prozesses und der Ergebnisse von Clusteranalysen. (Es wird gezeigt, was man mit Daten (so alles) tun kann, und was dabei herauskommt.) Dabei wird der für den Einsatz in einer empirischen Wissenschaft **zentrale Modellaspekt vernachlässigt**. Dieser betont, daß Ergebnisse von Repräsentationsverfahren Daten abzubilden haben, und nicht umgekehrt die Daten lediglich Ausgangspunkt von Verfahrensanwendungen sind.

Aus dem Interesse, den Modellaspekt der Resultate von Verfahrensanwendungen hervorzuheben, verstehe ich unter ‚Clusteranalyse‘ Methoden, die die Approximation einer Teilkasse (numerisch evaluierter) diskreter mathematischer Modelle an empirische Daten leisten.

Diese Sichtweise führt für die Entwicklung der methodologischen Literatur zu folgenden Forderungen:

- Da bei der Approximation von Modellen grundsätzlich deren Angemessenheit in Frage steht, sollten verstärkt vor allem **überprüfungsverfahren** entwickelt und dem potentiellen Benutzer auch in praktikabler Form zur Verfügung gestellt werden (Veröffentlichung/Austausch von Programmen; Implementation zumindest in die weit verbreiteten Pakete). Es ist nicht weiter akzeptabel, daß Programme zur Durchführung clusteranalytischer Verfahren gewöhnlich keinerlei Indizes zur Beurteilung der statistischen und/oder praktischen Signifikanz bzw. Bedeutsamkeit des Resultats mitteilen und eine Untersuchung von Residuen nicht vorgesehen ist (z.B. auch bei Programmen zur Faktorenanalyse).
- Die Approximation der diskreten Modelle sollte grundsätzlich als Problem der **Optimierung einer Zielfunktion** (unter Nebenbedingungen) angesehen

werden. (Dies könnte das ‚Gewirr‘ mathematischer Sprachen und die ‚Anhäufung‘ heuristischer Methoden abbauen helfen.) Solche Optimierungsprobleme werden bereits seit einiger Zeit in der mathematischen Verfahrensforschung (Operations Research) unter den Überschriften ‚Integer-Programmierung‘ und ‚diskrete‘, ‚kombinatorische‘, sowie ‚mathematische Programmierung‘ diskutiert. (Deshalb sei ausdrücklich auf die Bibliographien von Kasting (1976) und Hausmann (1978) und stellvertretend auf die Periodica ‚Computing‘ und ‚Mathematical Programming‘ hingewiesen.) Zwar gibt es einige Autoren, die diesen Ansatz auch im Bereich der Clusteranalyse verfolgt haben (z.B. Ward, 1963; Hartigan, 1967; Rao, 1971; Brucker, 1978; sowie Carroll und Hubert in verschiedenen Arbeiten), in der psychologischen mathematisch-statistischen Literatur hat sich aber die selbstverständliche Beachtung von Ansätzen zur Modellapproximation durch Optimierungsverfahren (siehe z.B. Horst, 1979; Dixon/Spedicato/Szegö, 1980) noch nicht allgemein durchgesetzt (zu ihrer Bedeutung in der Statistik siehe Kennedy/Gentle, 1980).

Das übliche **forschungspraktische Vorgehen** nach dem Schema: Datenerhebung/Datenerfassung/Programmanwendung (mit Standardoptionen)/Interpretation der Resultate sollte bei verstärkter Beachtung des Modellaspekts von Ergebnissen clusteranalytischer Repräsentationsverfahren durch einige Kontrollüberlegungen bereichert werden:

- Es ist die Frage zu beantworten, ob die Approximation eines diskreten Modells Gegenstand und Zielvorstellung der Untersuchung angemessen ist. (Für die HCA wird dieses Problem von Kleiter/Fillbrandt, 1973, behandelt.) Durch theoretisch inhaltliche Überlegungen ist also zu begründen, warum die Repräsentation von untersuchten Objekten, z.B. durch eine Mengenerlegung, adäquat sein soll. (Durch den Einsatz eines Verfahrens wird die Gruppierungsdarstellung erzwungen.)
- Die Geltung der Voraussetzungen von Verfahrensbestandteilen ist für vorliegende Daten zu prüfen (z.B. die Linearität der Zusammenhänge bei Verwendung von Korrelationen oder euklidischen Distanzen als Proximitäten).
- Die Gültigkeit der diskreten Modelle ist (soweit möglich) zumindest einer konfirmatorischen statistischen Prüfung zu unterziehen (s.o., z.B. Abschnitte **3.2** und **3.3**).
- Liegen zur Approximation von diskreten Modellen lediglich heuristische Verfahren (z.B. HCA) vor, so sind stets mehrere Lösungen zur Repräsentation der Daten zu generieren und mittels eines quantitativen Maßes der Anpassungsgüte zu bewerten.
- Schließlich sind die Abweichungen zwischen den Daten und der Modelldarstellung im einzelnen zu untersuchen, denn - wie John W. Tukey sagt -:

„Residuals are our main tool in going further. They are to the data analyst what powerful magnifying glasses, sensitive chemical tests for bloodstains, and delicate listening devices are to a story-book detective. They permeate all sorts of data analysis and appear in many guises.“ (1977, S. 125)

## Literatur

Diese Liste enthält nur Titel, die in der vorliegenden Arbeit zitiert wurden. - Bibliographischen Charakter haben die Literaturverzeichnisse in Bock (1974), Sneath/Sokal (1973) und Vogel (1975). Für neuere Arbeiten siehe ergänzend Eckes/Rosbach (1980).

- Abadie, J. (Ed). 1970. Integer and nonlinear programming. Amsterdam: North-Holland.
- Ahrens, H. J. 1974. Multidimensionale Skalierung. Weinheim: Beltz.
- Aldenderfer, M. S. & Blashfield, R. K. 1978. Computer Programms for performing hierarchical cluster analysis. *Applied Psychological Measurement*, 2, 405-413.
- Anderberg, M. R. 1973. Cluster analysis for applications. New York: Academic Press.
- Arabie, Ph. & Boorman, S. A. 1973. Multidimensional scaling of measures of distance between partitions. *Journal of Mathematical Psychology*, 10, 148-203.
- Arabie, Ph., Boorman, S. A. & Levitt, P. R. 1978. Constructing blockmodels: How and why. *Journal of Mathematical Psychology*, 17, 21-63.
- Arabie, Ph. & Carroll, J. D. 1980. MAPCLUS: A mathematical programming approach to fitting the ADCLUS model. *Psychometrika*, 45, 211-235.
- Atchley, W. R. & Bryant, E. H. (Eds). 1975. Multivariate statistical methods. Among-groups covariation. Stroudsburg: Dowden, Hutchinson & Ross, Inc.
- Baker, F. B. 1974. Stability of two hierarchical grouping techniques. Case 1: Sensivity to data errors. *Journal of the American Statistical Association*, 69, 440-445.
- Baker, F. B. & Hubert, L. J. 1975. Measuring the power of hierarchical cluster analysis. *Journal of the American Statistical Association*, 70, 31-38.
- Baker, F. B. & Hubert, L. J. 1976. A graph-theoretic approach to goodness-of-fit in complete-link hierarchical clustering. *Journal of the American Statistical Association*, 71, 870-878.
- Baker, F. B. & Hubert, L. J. 1977. Applications of combinatorial programming to data analyses: Seriation using asymmetric proximity measures. *British Journal of Mathematical and Statistical Psychology*, 30, 154-164.
- Balas, E. & Padberg, M. W. 1975. Set partitioning: A survey. In: Roy, B. (Ed.). 205-258.
- Banfield, C. F. & Basill, L. C. 1977. Algorithm AS 113: A transfer algorithm for non-hierarchical classification. *Applied Statistics*, 26, 206-210.
- Beale, E. M. L. 1970. Selecting an optimal subset. In: Abadie, J. (Ed.). 451-562.
- Beale, E. M. L., Kendall, M. G. & Mann, D. W. 1967. The discarding of variables in multivariate analysis. *Biometrika*, 54, 357-366.

- Berge, C. 1973, 1976<sup>2</sup>. Graphs and hypergraphs. Amsterdam: North-Holland.
- Blashfield, R. K. 1976. Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchical methods. *Psychological Bulletin*, 83, 377-388.
- Blashfield, R. K. 1980. The growth of cluster analysis: Tryon, Ward, and Johnson. *Multivariate Behavioral Research*, 15, 439-458.
- Blashfield, R. K. & Aldenderfer, M. S. 1978. The literature on cluster analysis. *Multivariate Behavioral Research*, 13, 271-295.
- Blashfield, R. K. & Aldenderfer, M. S. 1978. Computer Programms for performing iterative partitioning cluster analysis. *Applied Psychological Measurement*, 2, 533-541.
- Block, H. G. & Böck, G. et al. 1974. Klassifizierungsprozesse bei der Auswertung von Röntgenaufnahmen. In: Klix, F., Sydow, H. & Wysotzki, F. (Eds). 157-178.
- Bock, H. H. 1974. Automatische Klassifikation. Göttingen: Vandenhoeck & Ruprecht.
- Bonner, R. E. 1964. On some clustering techniques. *IBM Journal of Research and Development*, 8, 22-32.
- Boorman, S. A. & Arabie, Ph. 1972. Structural measures and the method of sorting: In: Shepard, R. N., Romney, A. K. & Nerlove, S. B. (Eds). 225-249.
- Boorman, S. A. & Olivier, D. C. 1973. Metrics on spaces of finite trees. *Journal of Mathematical Psychology*, 10, 26-59.
- Boyce, D. E., Farhi, A. & Weischedel, R. 1974. Optimal subset selection. Multiple regression, interdependence and optimal network algorithms. Berlin: Springer.
- Boyd, J. P. & Wexler, K. N. 1973. Trees with structure. *Journal of Mathematical Psychology*, 10, 115-147.
- Bron, C. & Kerbasch, J. 1973. Algorithm 457. Finding all cliques of an undirected graph. *Communications of the ACM*, 16, 575-577.
- Brucker, P. 1978. On the complexity of clustering problems. In: Henn, R., Korte, B., & Oettli, W., 45-54.
- Büttner, J. 1975. Zur Clusteranalyse. *Biometrical Journal*, 17, 163-179.
- Bunemann, P. 1971. The recovery of trees from measures of dissimilarity. In: Hodson, F. R., Kendall, D. G. & Tautu, P. (Eds). 387-395.
- Carroll, J. D. 1976. Spatial, non-spatial and hybrid models for scaling. *Psychometrika*, 41, 439-463.
- Carroll, J. D. & Chang, J. J. 1973. A method for fitting a class of hierarchical tree structure models to dissimilarities and its application to some „Body Parts“ data of Miller's. *Proceedings of the 81st Annual Convention of the American Psychological Association*, 8, 1097-1098.
- Garroll, J. D. & Pruzansky, S. 1980. Discrete and hybrid scaling models. In: Lantermann, E. D. & Feger, H. (Eds). 108-139.
- Christofides, N. 1975. Graph theory. An algorithmic approach. London: Academic Press.

- Christofides, N. 1975. Hamiltonian circuits and the travelling salesman problem: A survey. In: Roy, B. (Ed.). 149-171.
- Cohen, J. 1978. Multiple regression as a general data analytic system. *Psychological Bulletin*, 70, 426-433.
- Cole, A. J. (Ed.). 1969. Numerical taxonomy. London: Academic Press.
- Cole, A. J. & Wishart, D. 1970. An improved algorithm for the Jardine-Sibson method of generating overlapping clusters. *Computer Journal*, 13, 156-163.
- Colonius, H. & Schulze, H.-H. 1977. Trees constructed from empirical relations. *Braunschweiger Berichte aus dem Institut für Psychologie* 1.
- Colonius, H. & Schulze, H.-H. 1979. Repräsentation nichtnumerischer Ähnlichkeitsdaten durch Baumstrukturen. *Psychologische Beiträge*, 21, 98-111.
- Cormack, R. M. 1971. A review of classification. *Journal of the Royal Statistical Society (Series A)* 134, 321-367.
- Corsten, L. C. A. & Hermans, J. (Eds). 1978. *Compstat 1978. Proceedings in computational statistics*. Wien: Physica.
- Cunningham, J. P. 1974. Finding the optimal tree-realization of a proximity matrix. Paper presented at the Mathematical Psychology Meetings, Ann Arbor, August 1974.
- Cunningham, J. P. 1978. Free trees and bidirectional trees as representations of psychological distances. *Journal of Mathematical Psychology*, 17, 165-188.
- Cunningham, K. M. & Ogilvie, J. C. 1972. Evaluation of hierarchical grouping techniques: A preliminary study. *Computer Journal*, 15, 209-213.
- D'Andrade, R. 1978. U-statistic hierarchical clustering. *Psychometrika*, 43, 59-67.
- Day, W. H. E. 1977. Validity of clusters formed by graph-theoretic cluster methods. *Mathematical Biosciences*, 36, 299-317.
- Degerman, R. 1970. Multidimensional analysis of complex structure: Mixtures of class and quantitative Variation. *Psychometrika*, 35, 475-491.
- Degerman, R. 1972. The geometric representation of some simple structures. In: Shepard, R. N., Romney, A. K. & Nerlove, S. B. (Eds). 194-211.
- Dixon, L. C. W., Spedicato, E. & Szegö, G. P. (Eds). 1980. *Nonlinear optimization. Theory and algorithms*. Boston: Birkhäuser.
- Dixon, W. J. (Ed.). 1975. *BMDP. Biomedical computer programs*. Berkeley: University of California Press.
- Dixon, W. J. & Brown, M. B. 1977. *BMDP-77. Biomedical Computer programs. P-Series*. Berkeley: University of California Press, 418ff.
- Duran, B. S. & Odell, P. L. 1974. Cluster analysis. Berlin: Springer.
- Eckes, Th. & Roßbach, H. 1980. *Clusteranalysen*. Stuttgart: Kohlhammer.
- Edelbrock, C. 1979. Mixture model tests of hierarchical clustering algorithms: The problem of classifying everybody. *Multivariate Behavioral Research*, 14, 367-384.

- Edelbrock, C. & McLaughlin, B. 1980. Hierarchical cluster analysis using intraclass correlations: A mixture model study. *Multivariate Behavioral Research*, 15, 299-384.
- Edgington, E. S. 1969. Statistical inference: The distribution-free approach. New York: Mc Graw-Hill.
- Edginton, E. S. 1980. Randomization tests. New York: Marcel Dekker.
- Edwards, A. W. F. & Cavalli-Sforza, L. L. 1965. A method for cluster analysis. *Biometrics*, 21, 362-375.
- Eisler, H. 1973. The algebraic and statistical tractability of the city block metric. *British Journal of Mathematical and Statistical Psychology*, 26, 212-218.
- Enslein, K., Raiston, A. & Wilf, H. S. (Eds). 1977. Statistical methods for digital Computers (Vol. 3). New York: Wiley.
- Estabrook, G. F. (Ed.). 1975. Proceedings of the Eighth International Conference on Numerical Taxonomy. S. F.: Freeman.
- Eye, A. von 1977a. über die Verwendung von Quadriken zur einbeschreibenden Klassifikation. *Biometrical Journal*, 19, 283-290.
- Eye, A. von 1977b. Zum Vergleich zwischen der hierarchischen Clusteranalyse nach Ward und MACS, einer mehrdimensionalen automatischen Clustersuchstrategie. *Psychologische Beiträge*, 19, 201-217.
- Eye, A. von & Wirsing, M. 1978. An attempt for a mathematical foundation and evaluation of Macs, a method for multidimensional automatical cluster detection. *Biometrical Journal*, 20, 655-666.
- Farris, J. S. 1969. On the cophenetic correlation coefficient. *Systematic Zoology* 18, 279-285. In: Atchley, W. R. & Bryant, E. H. (Eds). 1975, 398-404.
- Fisher, W. D. 1958. On grouping for maximum homogeneity. *Journal of the American Statistical Association*, 53, 789-798.
- Flachsmeier, J. 1977<sup>3</sup>. Kombinatorik. Berlin (Ost): VEB Deutscher Verlag der Wissenschaften.
- Frank, O. 1978. Inferences concerning cluster structure. In: Corsten, L. C. A. & Hermans, J. (Eds). 259-265.
- Friendly, M. 1977. In search of the M-Gram: The structure of organization in free recall. *Cognitive Psychology*, 9, 188-249.
- Furnival, G. M. & Wilson, R. W. 1974. Regression by leaps and bounds. *Technometrics*, 16, 499-511.
- Gower, J. C. 1975. Goodness-of-fit criteria for classification models and other patterned structures. In: Estabrook, G. F. (Ed.). 38-62.
- Gower, J. C. & Ross, G. J. S. 1969. Minimum spanning trees and single linkage cluster analysis. *Applied Statistics*, 18, 54-64.
- Harary, F. 1974. Graphentheorie. München: Oldenbourg.
- Hartigan, J. A. 1967. Representation of similarity matrices by trees. *Journal of the American Statistical Association*, 62, 1140-1158.

- Hartigan, J. A. 1972. Direct clustering of a data matrix. *Journal of the American Statistical Association*, 67, 123-129.
- Hartigan, J. A. 1975: Clustering algorithmus. New York: Wiley.
- Hartigan, J. A. 1977. Distribution problems in clustering. In: van Ryzin, J. (Ed.). 45 - 71.
- Hausmann, D. 1978. Integer programming and related areas. A classified bibliography, 1976-1978. Berlin: Springer.
- Henn, R., Korte, B. & Oettli, W. 1978. Optimization and Operations Research. Proceedings, Bonn 1977. Berlin: Springer.
- Herden, W. & Steinhausen, D. 1979. Ein Verfahren zur Berechnung eines globalen Minimums beim Varianzkriterium in der Cluster-Analyse. In: Späth, H. (Ed.). 115-142.
- Hodson, F. R., Kendall, D. G. & Tautu, P. 1971. Mathematics in the archaeological and historical sciences. Edinburgh University Press.
- Horst, R. 1979. Nichtlineare Optimierung. München: Hanser.
- Hubert, L. J. 1972. Some extensions of Johnson's hierarchical clustering algorithms. *Psychometrika*, 37, 261-274.
- Hubert, L. J. 1973a. Monotone invariant clustering procedures. *Psychometrika*, 38, 47-62.
- Hubert, L. J. 1973b. Min and max hierarchical clustering using asymmetric similarity measures. *Psychometrika*, 38, 63-72.
- Hubert, L. J. 1974a. Approximative evaluation techniques for the single-link and complete-link hierarchical clustering procedures. *Journal of the American Statistical Association*, 69, 698-704.
- Hubert, L. J. 1974b. Some applications of graph theory to clustering. *Psychometrika*, 39, 283-309.
- Hubert, L. J. 1974c. Spanning trees and aspects of clustering. *British Journal of Mathematical and Statistical Psychology*, 27, 19-88.
- Hubert, L. J. 1976. Seriation using asymmetric proximity measures. *British Journal of Mathematical and Statistical Psychology*, 29, 32-52.
- Hubert, L. J. 1977a. A set-theoretical approach to the problem of hierarchical clustering. *Journal of Mathematical Psychology*, 15, 70-88.
- Hubert, L. J. 1977b. Data analysis implications of some concepts related to the cuts of a graph. *Journal of Mathematical Psychology*, 15, 199-208.
- Hubert, L. J. 1978. Generalized proximity function comparison. *British Journal of Mathematical and Statistical Psychology*, 31, 179-192.
- Hubert, L. J. 1979a. Comparison of sequences. *Psychological Bulletin*, 86, 1098-1106.
- Hubert, L. J. 1979b. Matching models in the analysis of cross-classifications. *Psychometrika*, 44, 21-41.

- Hubert, L. J. 1979c. Generalized concordance. *Psychometrika*, 44, 135-142.
- Hubert, L. J. 1980. Analyzing proximity matrices: The assessment of internal variation in combinatorial structure. *Journal of Mathematical Psychology*, 21, 247-264.
- Hubert, L. J. & Baker, F. B. 1977a. An empirical comparison of baseline models for goodness-of-fit in r-diameter hierarchical clustering. In: Ryzin, J. van (Ed.). 131-153.
- Hubert, L. J. & Baker, F. B. 1977b. The comparison and fitting of given classification schemes. *Journal of Mathematical Psychology*, 16, 233-253.
- Hubert, L. J. & Baker, F. B. 1978a. Analyzing the multitrait-multimethod matrix. *Multivariate Behavioral Research*, 13, 163-179.
- Hubert, L. J. & Baker, F. B. 1978b. Applications of combinatorial programming to data analysis: The traveling salesman and related problems. *Psychometrika*, 43, 81-91.
- Hubert, L. J. & Baker, F. B. 1979. A note on analyzing the multitrait-multimethod matrix: An application of a generalized proximity function comparison. *British Journal of Mathematical and Statistical Psychology*, 32, 179-184.
- Hubert, L. J. & Schultz, J. V. 1975. Hierarchical clustering and the concept of space distortion. *British Journal of Mathematical and Statistical Psychology*, 28, 121-133.
- Hubert, L. J. & Schultz, J. 1976. Quadratic assignment as a general data analysis strategy. *British Journal of Mathematical and Statistical Psychology*, 29, 190-241.
- Hull, C. H. & Nie, N. N. 1979. SPSS-Update: New procedures and facilities for releases 7 and 8. New York: Mc Graw-Hill.
- Jardine, N. & Sibson, R. 1968. The construction of hierarchic and non-hierarchic classifications. *Computer Journal*, 11, 177-184.
- Jardine, N. & Sibson, R. 1971. Mathematical taxonomy. New York: Wiley.
- Jöreskog, K. G. 1978. Structural analysis of covariance and correlation matrices. *Psychometrika*, 43, 443-477.
- Johnson, S. C. 1967. Hierarchical clustering schemes. *Psychometrika*, 32, 241-254.
- Kallina, H. 1967. Das Unbehagen in der Faktorenanalyse. *Psychologische Beiträge*, 10, 81-86.
- Kasting, C. 1976. Integer programming and related areas. A classified bibliography. Berlin: Springer.
- Kaufmann, A. 1975. Introduction to the Theory of fuzzy subsets. New York: Academic Press.
- Kennedy, W. J. & Gentle, J. E. 1980. Statistical computing. New York: Marcel Dekker Inc.
- Kirn, Ki Hang & Roush, F. W. 1978. Ultrametrics and matrix theory. *Journal of Mathematical Psychology*, 18, 195-203.

- Kleiter, E. F. 1980. Voraussetzungs-Koeffizient und Voraussetzungs-Clusteranalyse bei numerischen Daten. Teil 1: Doppelte Abbildung und Voraussetzungs-Clusteranalyse. Teil II: Koeffizienten und Anwendungsbeispiele. Zeitschrift für erziehungswissenschaftliche Forschung, 14, 65-89 und 127-151.
- Kleiter, E. F. & Fillbrandt, H. 1973. Hypothesenorientierte Kategorienbildung und hierarchische Cluster-Analyse. Psychologische Beiträge, 15, 603-633.
- Kleiter, E. F. & Petermann, F. 1977. Abbildung von Lernwegen. München: Oldenbourg.
- Kleiter, E. F. & Timmermann, G. 1977. über einen Algorithmus zur hierarchischen Voraussetzungs-Struktur-Analyse. Psychologische Beiträge, 19, 355-390.
- Klix, F., Sydow, H. & Wysotzki, F. (Eds). 1974. Erkennungs- und Klassifizierungsprozesse. Berlin (Ost): VEB Verlag der Wissenschaft.
- Kopp, B. 1978a. Hierarchical classification I: Single-linkage method. Biometrical Journal, 20, 495-501.
- Kopp, B. 1978b. Hierarchical classification II: Complete-linkage method. Biometrical Journal, 20, 597-602.
- Kopp, B. 1978c. Hierarchical classification III: Average-linkage, median, centroid, WARD, flexible Strategie. Biometrical Journal, 20, 703-711.
- Krishnaiah, P. R. (Ed.). 1969. Multivariate analysis Vol. II. New York: Academic Press.
- Kruskal, J. 1977. The relationship between multidimensional scaling and clustering. In: Ryzin, J. van (Ed.). 17-44.
- Kruskal, J. & Carroll, J. B. 1969. Geometrical models and badness-of-fit functions. In: Krishnaiah, P. R. (Ed.). 639-671.
- Kubicki, St., Herrmann, W. M. & Laudahn, G. (Eds). 1980. Faktorenanalyse und Variablenbildung aus dem Elektroenzephalogramm. Stuttgart: Fischer.
- Kuiper, F. K. & Fisher, L. 1975. A Monte Carlo comparison of six clustering procedures. Biometrics, 31, 777-783.
- Lance, G. N. & Williams, W. T. 1967. A general theory of classificatory sorting strategies: I. Hierarchical Systems. Computer Journal, 9, 373-380.
- Lantermann, E. D. & Feger, H. (Eds). 1980. Similarity and choice. Papers in honour of Clyde Coombs. Bern: Huber.
- Lawler, E. L. 1975. The quadratic assignment problem: A brief review. In: Roy, B. (Ed.). 351-360.
- Lazarsfeld, P. F. & Henry, N. W. 1968. Latent structure analysis. Boston: Houghton, Mifflin.
- Lee, H. B. & Mac Queen, J. B. 1980. A K-means cluster analysis Computer programm with cross-tabulations and next-nearest-neighbor analysis. Educational and Psychological Measurement, 40, 130-138.
- Leuschner, D. 1973. Numerische Taxonomie und algebraische Struktur. Biometrical Journal, 15, 271-275.

- Lienert, G. A. 1978. Verteilungsfreie Methoden in der Biostatistik, Band II. Meisenheim: Hain.
- Ling, R. F. 1971. Cluster analysis. New Haven, Conn.: Dissertation, Yale University.
- Ling, R. F. 1972. On the theory and construction of K-clusters. Computer Journal, 15, 326-332.
- Ling, R. F. 1973. A probability theory of cluster analysis. Journal of the American Statistical Association, 68, 159-164.
- Ling, R. F. 1975. An exact probability distribution on the connectivity of random graphs. Journal of Mathematical Psychology, 12, 90-98.
- Ling, R. F. & Killough, G. G. 1976. Probability tables for cluster analysis based on a theory of random graphs. Journal of the American Statistical Association, 71, 293-300.
- Lingoes, J. C. 1973. The Guttman-Lingoes nonmetric program series. Ann Arbor: Mathesis Press.
- Lingoes, J. C. & Cooper, T. 1971. PEP-I: A Fortran IV (G) program for Guttman-Lingoes nonmetric probability clustering. Behavioral Science, 16, 259-261.
- Lingoes, J. C., Roskam, E. E. & Borg, J. 1976. Geometric representations of relational data structures. Ann Arbor: Mathesis Press.
- Marsten, R. E. 1975. An algorithm for large set partitioning problems. In: Roy, B. (Ed.). 259-267.
- Matejcek, M. & Schenk, G. K. 1976. Quantitative analysis of the EEG. Proceedings of 2nd Symposium of the Study Group for EEG Methodology, Jongny sur Vevey, Mai 1975; Konstanz: AEG-Telefunken.
- Matula, D. W. 1977. Graph theory techniques for cluster analysis algorithms. In: Ryzin, J. van (Ed.). 95-129.
- Mc Henry, C. E. 1978. Computation of a best subset in multivariate analysis. Applied Statistics, 27, 291-296.
- Mc Intyre, R. M & Blashfield, R. K. 1980. A nearest-centroid technique for evaluating the minimum-variance clustering procedure. Multivariate Behavioral Research, 15, 225-238.
- Mc Quitty, L. L. 1968. Multiple clusters, types, and dimensions from iterative intercolumnar correlational analysis. Multivariate Behavioral Research, 3, 465-477.
- Mc Quitty, L. L. & Clark, J. A. 1968. Clusters from iterative, intercolumnar correlational analysis. Educational and Psychological Measurement, 28.
- Milligan, G. W. 1979. Ultrametric hierarchical clustering algorithms. Psychometrika, 44, 343-346.
- Milligan, G. W. 1980. An examination of the effect of six types of error perturbation on fifteen clustering algorithms. Psychometrika, 45, 325-342.
- Milligan, G. W. & Sokol, L. M. 1980. A two-stage clustering algorithm with robust recovery characteristics. Educational and Psychological Measurement, 40, 755-759.

- Mustonen, S. 1978. Digression analysis: Fitting alternative regression models to heterogenous data. In: Corsten, L. C. A. & Hermans, J. (Eds). 95-101.
- Müller-Merbach, H. 1970. Optimale Reihenfolgen. Berlin: Springer.
- Negoita, C. V. & Ralescu, D. A. 1975. Applications of fuzzy sets to systems analysis. Basel: Birkhäuser.
- Nenninger, P. 1980. Anwendungsmöglichkeiten der Graphentheorie in der Erziehungswissenschaft. Zeitschrift für Empirische Pädagogik, 4, 85-106.
- Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K. & Bent, D. H. 1975<sup>2</sup>. SPSS - Statistical package for the social sciences. New York: Mc Graw-Hill.
- Oldenbürger, H.-A. & Becker, D. 1976. Are there clusters of frequencies in power-spectra of EEG? How to find and prove them statistically. In: Matejcek, M. & Schenk, G. K. (Eds). 601-611.
- Oldenbürger, H.-A. & Schwibbe, M. 1980. Konstruktive Kritik des Einsatzes dimensionaler Dekompositionsverfahren für EEG-Frequenzkomponenten. In: Kubicki, St., Herrmann, W. M. & Laudahn, G. (Eds). 47-60.
- Opitz, O. 1980. Numerische Taxonomie. Stuttgart: Fischer, UTB 918.
- Patrinos, A. N. & Hakimi, S. L. 1972. The distance matrix of a graph and its tree realization. Quarterly of Applied Mathematics, 30, 255-269.
- Peay, E. R. 1975. Nonmetric grouping: Clusters and cliques. Psychometrika, 40, 297-313.
- Rand, W. M. 1971. Objective criteria for the evaluation of clustering methods. Journal of the American Statistical Association, 66, 846-850.
- Rao, M. R. 1971. Cluster analysis and mathematical programming. Journal of the American Statistical Association, 66, 622-626.
- Revelle, W. 1978. INCLUST: A cluster analytic approach to exploratory and confirmatory scale construction. Behavioral Research Methods and Instrumentation, 10, 739-742.
- Revelle, W. 1979. Hierarchical cluster analysis and the internal structure of tests. Multivariate Behavioral Research, 14, 57-74.
- Rohlf, F. G. 1974. Graphs implied by the Jardine-Sibson overlapping clustering methods,  $B_k$ . Journal of the American Statistical Association, 69, 705-710.
- Rohlf, F. J. 1975. A new approach to the computation of the Jardine-Sibson  $B_k$  clusters. Computer Journal, 18, 164-168.
- Rasch, E. 1975. Cognitive reference points. Cognitive Psychology, 7, 532-547.
- Rasch, E. 1978. Principles of categorization. In: Rasch, E. & Lloyd, B. B. (Eds). 27-48.
- Rasch, E. & Lloyd, B. B. (Eds). 1978. Cognition and categorization. Hillsdale, N. J.: Erlbaum.
- Roskam, E. E. 1975. Nonmetric data analysis. Department of Psychology, University of Nijmegen, Holland, Report 75-MA-13.

- Roy, B. (Ed.). 1975. Combinatorial programming: Methods and applications. Proceedings of the NATO Advanced Study. Institut Dordrecht (Holland); D. Reidel Publ. Comp., 1975.
- Ryzin, J. van (Ed.). 1977. Classification and clustering. New York: Academic Press.
- Sattath, S. & Tversky, A. 1977. Additive similarity trees. *Psychometrika*, 42, 319-345.
- Schultz, J. V. & Hubert, L. 1973. Data analysis and the connectivity of random graphs. *Journal of Mathematical Psychology*, 10, 421-428.
- Shepard, R. N. & Arabie, Ph. 1979. Additive clustering: Representation of similarities as combinations of discrete overlapping properties. *Psychological Review*, 86, 87-123.
- Shepard, R. N., Romney, A. K. & Nerlove, S. B. (Eds). 1972. Multidimensional scaling. Theory and application in the behavioral sciences. Vol. 1. New York: Seminar Press.
- Silverstein, A. B. 1980. Cluster analysis of the Wechsler intelligence scale for children-revised. *Educational and Psychological Measurement*, 40, 51-54.
- Sneath, P. H. A. 1969. Evaluation of clustering methods. In: Cole, A. J. (Ed.). 257-271.
- Sneath, P. H. A. & Sokal, R. R. 1973. Numerical taxonomy. S. F.: Freeman.
- Sodeur, W. 1976. Buchbesprechung zu Vogel 1975. *Allgemeines Statistisches Archiv*, 415-416.
- Sokal, R. R. & Rohlf, F. J. 1962. The comparison of dendograms by objective methods. *Taxon*, 11, 33-39.
- Sokal, R. R. & Sneath, P. H. A. 1963. Principles of numerical taxonomy. S. F.: Freeman.
- Späth, H. 1975. Cluster-Analyse-Algorithmen zur Objektklassifizierung und Datenreduktion. München: Oldenbourg.
- Späth, H. 1976.  $L_1$  Cluster Analyse. *Computing*, 16, 379-387.
- Späth, H. 1977. Fallstudien Cluster-Analyse. München: Oldenbourg.
- Späth, H. 1977a. Partitionierende Cluster-Analyse für große Objektmengen mit binären Merkmalen am Beispiel von Firmen und deren Berufsgruppenbedarf. In: Späth, H., 1977, 63-80.
- Späth, H. 1977b. Computational experiences with the exchange method applied to four commonly used partitioning cluster analysis criteria. *European Journal of Operation Research*, 1, 23-31.
- Späth, H. (Ed.). 1979. Ausgewählte Operations Research Software in FORTRAN. München: Oldenbourg.
- Späth, H. & Müller, R. 1979. Das Austauschverfahren für die skalierungsvariante Methode der adaptiven Distanzen in der Cluster-Analyse. In: Späth, H. (Ed.). 143-163.

- Steiger, J. H. 1980: Testing pattern hypotheses on correlation matrices: Alternative statistics and some empirical results. *Multivariate Behavioral Research*, 15, 335-352.
- Steinhausen, D. & Langer, K. 1977. Clusteranalyse. Einführung in Methoden und Verfahren der automatischen Klassifikation. Berlin: de Gruyter.
- Tinhofer, G. 1980. Zufallsgraphen. München: Hanser.
- Tyron, R. C. 1939. Cluster analysis. Ann Abour, Mich.: Edwards Brothers.
- Tukey, J. W. 1977. Exploratory data analysis. Reading, Mass.: Addison-Wesley.
- Tversky, A. 1977. Features of similarity. *Psychological Review*, 84, 327-352; auch in: Lantermann, E. D. & Feger, H. (Eds). 1980, 12-53.
- Tversky, A. & Sattath, S. 1979. Preference trees. *Psychological Review*, 86, 542-573.
- Tzeng, O. C. S. & May, W. H. 1979. On rotation of a Johnson hierarchical tree structure. *Educational and Psychological Measurement*, 39, 733-741.
- Vogel, F. 1975. Probleme und Verfahren der numerischen Klassifikation. Göttingen: Vandenhoeck & Ruprecht.
- Ward, J. H. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58, 236-244.
- White, H. C. 1977. Probabilities of homomorphic mappings from multiple graphs. *Journal of Mathematical Psychology*, 16, 121-134.
- Williams, W. T. & Lance, G. N. 1977. Hierarchical classificatory methods. In: Enslein, K., Ralston, A. & Wilf, H. S. (Eds). 269-295.
- Wishart, D. 1978. CLUSTAN. User manual, Program library unit, Edinburgh University. (Program Library Unit, Edinburgh University 18 Bucclech Place, Edinburgh EH8 9LN Scotland).
- Welfe, J. H. 1970. Pattern clustering by multivariate mixture analysis. *Multivariate Behavioral Research*, 5, 329-350.
- Wolfe, J. H. 1978. Comparative cluster analysis of patterns of vocational interest. *Multivariate Behavioral Research*, 13, 33-44.
- Woodward, J. A. & Bentler, P. M. 1979. Application of optimal sign-vectors to reliability and cluster analysis. *Psychometrika*, 44, 337-342.
- Wright, W. E. 1974. An axiomatic specification of euclidean analysis. *Computer Journal*, 17, 355-364.
- Yahil, A. & Brown, M. B. 1976. On separating clusters from background. *Technometrics*, 18, 55-58.
- Yeh, R. T. & Bang, S. Y. 1975. Fuzzy relations, fuzzy graphs, and their applications to clustering analysis. In: Zadeh, L. A., Fu, King-Sun, Tanaka, K. & Shimura, M. (Eds). 125-149.
- Young, F. W., de Leeuw, J. & Takane, Y. 1980. Quantifying qualitative data. In: Lantermann, E. D. & Feger, H. (Eds). 150-179.

- Zadeh, L. A. 1977. Fuzzy sets and their application to pattern classification and clustering analysis. In: van Ryzin, J. (Ed.). 251-299.
- Zadeh, L. A., Fu, King-Sun, Tanaka, K. & Shimura, M. (Eds). 1975. Fuzzy sets and their applications to cognitive and decision Prozesses. New York: Academic Press.

Lawrence Hubert danke ich für die Durchsicht der Druckvorlage dieses Artikels und seinen ergänzenden Hinweis auf das folgende wichtige Review, das leider nicht mehr eingearbeitet werden konnte:

- Dubes, R. & Jain, A. K. 1980. Clustering Methodologies in Exploratory Data Analysis. In: Yovits, M. C. Advances in Computers, Vol. 19, New York: Academic Press, 113-228.

## 8. Kapitel

# Typenanalyse

*Joachim Krauth*

### 1. Einführung

Eine Schwierigkeit bei der Darstellung der Verfahren, die man unter dem Oberbegriff *Typenanalyse* zusammenfaßt, besteht in der Abgrenzung zu Verfahren, die entweder von einzelnen Autoren oder auch von einem allgemeinen Standpunkt aus zur Typenanalyse gerechnet werden und deshalb hier abgehandelt werden müßten, jedoch schon an anderer Stelle der Enzyklopädie in eigenen Kapiteln besprochen werden. Um Überschneidungen weitgehend zu vermeiden, soll auf diese Gebiete hier nur kurz verwiesen werden und allenfalls bei Vergleich mit anderen Verfahren darauf eingegangen werden. So wird in Kapitel 3 (Band 2) das gelegentlich als Typenanalyse (vgl. Dorsch, 1970) bezeichnete *Q-sort-Verfahren* besprochen, in Kapitel 5 (Band 4) die *Diskriminanzanalyse*, in Kapitel 7 (Band 4) die *Clusteranalyse* und in Kapitel 6 (Band 4) die *Latente Strukturanalyse*. Alle diese Verfahren sind schon als Typenanalysen bezeichnet oder zur Aufstellung von Typen verwendet worden. Auch in früheren Handbuchbeiträgen sind typenanalytische Verfahren entweder direkt (Mittenecker, 1960; Mombour, 1977) oder unter dem Gesichtspunkt der Klassifikation angesprochen worden (Janke, 1971).

Auf die Definition von Typen gehen die Handbuchartikel von Graumann (1960), Strunz (1960) und Mombour (1977) ein, während Schick (1971), Dietsch & Volk (1977), Werner (1966), Rausch (1966), Graumann (1966), Kainz (1964), Kaminski (1964) und Groffmann & Schmidtke (1977) Typologien aus verschiedenen Bereichen der Psychologie beschreiben. Auf die Vielzahl der vorhandenen psychologischen Typologien und Definitionen für den Begriff „Typ“ kann hier nicht eingegangen werden. Während jedes sinnvolle typenanalytische Verfahren von einer exakten Definition des zugrundegelegten Typenbegriffs ausgehen muß, existieren umgekehrt viele Typologien, die auf rein heuristischer Basis gewonnen wurden, ohne methodisch hinreichend abgesichert zu sein. Weder die Typologien, die rein heuristisch, noch die Typologien, die mit Hilfe eines der oben ausgeklammerten Verfahren erhalten

wurden, sollen hier diskutiert werden. Für die verbleibenden typenanalytischen Verfahren werden jedoch jeweils die Definition eines Typs und etwaige mit Hilfe der Verfahren aufgestellte Typologien angegeben.

Der Begriff *Typenanalyse* wird in der Literatur durchaus unterschiedlich verstanden. In Baumann (1971, S. 23) wird er mit dem Begriff *Taxonomie* und sogar mit dem Begriff *Typologie* gleichgesetzt, und es wird darunter „die Erörterung der theoretischen Grundlagen und die praktische Durchführung der Gruppierung der Elemente“ verstanden. Dagegen wird „die Anwendung mathematisch-statistischer Verfahren auf Probleme der Taxonomie“ als *Taxometrie* bezeichnet.

Hofstätter (1940, 1955, 1957, 1971) versteht unter Typenanalyse im wesentlichen die sogenannte *Q-Analyse*, d.h. eine Faktorenanalyse über eine Korrelationsmatrix zwischen den Individuen, ebenso wie Eysenck (1941), Burt (1937), Stephenson (1963a, 1936b, 1950, 1952, 1953) und andere. Mit der Q-Analyse darf nicht verwechselt werden die *Q'-Analyse* im Sinne von Cattell (1957) (bzw. *Q-Analyse* in der Bezeichnungsweise von Cattell & Coulter (1966)). Es handelt sich dabei um Clusterverfahren, die auf eine Ähnlichkeitsmatrix zwischen den Personen angewandt werden, und nicht um faktorenanalytische Verfahren.

Hier soll unter Typenanalyse sowohl Taxonomie als auch Taxometrie im obigen Sinne verstanden werden. Nach Baumann (1971, S. 21) existieren zwei verschiedene Methoden der Typenfindung: die Clusteranalyse und die Faktorenanalyse. Dabei werden in der Clusteranalyse Gruppen ähnlicher Elemente und in der Faktorenanalyse die zugrundeliegenden Dimensionen gesucht. Hier sollen faktorenanalytische Verfahren (insbesondere die Q-Analyse) ausgeklammert werden und Clusteranalyseverfahren nur berücksichtigt werden, wenn sie zum Auffinden von Typen im psychologischen Sinne konzipiert wurden. Die Übergänge zwischen Typenanalyse und Clusteranalyse sind aber so fließend, daß Überschneidungen unvermeidlich erscheinen.

Als übersichtsarbeiten zur Typenanalyse mit teilweise anderen Schwerpunkten stehen u.a. Baumann (1971), Gigerenzer (1977) und Zerssen (1973, 1977) zur Verfügung. In Sneath & Sokal (1973) wird der Schwerpunkt mehr auf Verfahren im biologisch-medizinischen Anwendungsbereich gelegt, während Bock (1974) und Duran & Odell (1974) Clusterverfahren nach mathematischen Gesichtspunkten referieren.

Neben der Abgrenzung zu anderen Verfahren bereitet eine Gliederung typenanalytischer Methoden erhebliche Schwierigkeiten. Dieses liegt daran, daß viele der auf diesem Gebiet arbeitenden Forscher ihre Verfahren immer wieder modifiziert und neuen Anwendungsmöglichkeiten angepaßt haben unter Verwendung alter eigener Ansätze und der Ansätze anderer Autoren. Deshalb

sollen hier sukzessiv Verfahren einzelner Autoren bzw. Autorengruppen vorgestellt werden, wobei, soweit sinnvoll, innerhalb dieser Abschnitte eine chronologische Gliederung vorgenommen wird. Als besondere Schwerpunkte werden dabei die Musteranalyseverfahren von McQuitty und die Konfigurationsfrequenzanalyse von Lienert ausgezeichnet. Angesichts der kaum noch zu übersehenden Vielzahl der Publikationen zu diesen beiden Ansätzen erscheint es gerechtfertigt, hier eine möglichst vollständige Darstellung anzustreben, auch wenn dabei gelegentliche Überschneidungen mit anderen Gebieten nicht zu vermeiden sind.

## *2. Die Musteranalyse von McQuitty*

Die Grundidee der Musteranalyse scheint auf Zubin (1934, 1936, 1937, 1938) zurückzugehen. Im Unterschied zur Faktorenanalyse, die viele Variablen auf wenige zugrundeliegende Komponenten zurückführt, führt die Musteranalyse nach Zubin (1936) zu einer Aufteilung der Gesamtstichprobe in mehrere Typen, die durch spezifische Muster oder Syndrome charakterisiert sind. In Zubin (1937) werden Antwortmuster für jeweils 3 von  $k$  binären Testitems betrachtet und ausgezählt, wie viele Personen jeweils das gleiche Muster aufweisen. Dieses geschieht für je eine Stichprobe von Normalen und Kranken. Da es auch nicht beantwortete Items gibt, erhält man  $3^3 = 27$  verschiedene Muster für eine Triade von Items. Dieses führt zu einer  $(2 \times 27)$ -Kontingenztafel, wobei die Zeilen den Diagnosen und die Spalten den Mustern entsprechen. Durch Dateninspektion werden diejenigen  $r$  Muster ausgewählt, die zwischen den Stichproben am besten differenzieren. Die restlichen Muster werden in einer Restklasse zusammengefaßt. Die Güte der Diskrimination wird durch das Chiquadrat in einer  $(2 \times (r+1))$ -Tafel gemessen. Um echte Antworttypen von Zufallsmustern zu unterscheiden, bezeichnet Zubin nur solche Muster als Typen, die sich nicht auch bei völliger Unabhängigkeit der Items ergeben würden.

Eine Weiterentwicklung des Verfahrens wird in Zubin (1938) beschrieben. Ausgangspunkt sind  $N$  Individuen, die einen Fragebogen mit  $k$  binären Items auszufüllen haben. In eine  $(NxN)$ -Tafel wird für jedes Paar von Individuen die Anzahl der gleich beantworteten Items eingetragen. Falls die Reihenfolge der Individuen für die Zeilen und Spalten dieselbe ist, ergibt sich eine bezüglich der Hauptdiagonalen symmetrische Ähnlichkeitstafel. Die Elemente dieser Tafel heißen *Übereinstimmungsscores*. Die Elemente der Hauptdiagonalen sind dabei alle gleich der totalen Anzahl  $k$  der Items. Zu einer Stichprobe von  $N$  kranken Individuen wird eine Stichprobe von  $N$  gesunden Individuen gesucht, die bezüglich Alter, Geschlecht, Bildungsstand usw. parallelisiert ist. Dieses führt zu zwei Ähnlichkeitstafeln für kranke und gesunde Individuen. Für jede Stichprobe wird jedem Individuum die Anzahl der Gruppenmitglie-

der zugeordnet, mit denen es in mindestens  $k_1 < k$  Items übereinstimmt. Alle diese Personen zusammen bilden einen Typ. Dabei ist  $k_1$  eine willkürliche Schranke. Durch Betrachten eines  $k_2 < k_1$  erhält man einen Typ für den Rest der Stichprobe. Durch weitere schrittweise Erniedrigung der Übereinstimmungsschranke ergeben sich weitere Typen bis auf eine Restklasse. Als nächstes wird für jeden Typ durch Dateninspektion nach einem gemeinsamen Reaktionsmuster gesucht, in dem möglichst viele Personen eines Typs übereinstimmen. Durch unterschiedliche Muster bei Kranken und Gesunden ergibt sich die Möglichkeit, ein Individuum zu klassifizieren.

Einen Überblick über die frühen Versuche einer Musteranalyse geben Gaier und Lee (1953). Sie unterscheiden die Probleme der *Musterdarstellung* (Skalierung der zugrundeliegenden Variablen), der *Musterähnlichkeit* und der *Mustervorhersage*. In einer Gegenüberstellung werden der Ähnlichkeitskoeffizient von Zubin (1937), die Steigungsmethode von DuMas (1946, 1949, 1950), der Musterähnlichkeitskoeffizient  $r_p$  von Cattell (1949), der D-Koeffizient von Cronbach und Gleser (1952) bzw. Osgood und Suci (1952), die Checklist-Methode nach Cronbach (1950), die Quotientenmethode für Profilvariablen nach Cronbach (1949b), die Korrelationsbildung zwischen den Scores einschließlich Eta-Koeffizient und Q-Technik, die Mustertabellierung nach Cronbach (1949a, 1949b), sowie die multiple Regression und die Diskriminanzanalyse als Verfahren zum Vergleich von Testmustern miteinander verglichen. Getrennt von den Verfahren, die zur Analyse von Testprofilen verwendet werden können, diskutieren die Autoren Verfahren, die auf Reaktionsmuster für einzelne Tests angewandt werden. Hier werden vor allem der Ähnlichkeitskoeffizient von Zubin (1936, 1937) und das Vorgehen von McQuitty (1953) verglichen.

In Gaier, Lee & McQuitty (1953) wird die von McQuitty (1953) entwickelte *Konfigurationsanalyse* verwendet, um charakteristische Stile des logischen Denkens zu identifizieren. Dieses Verfahren misst den Assoziationsgrad zwischen Paaren von Reaktionsalternativen durch Phi-Koeffizienten und wendet auf die Korrelationsmatrix eine Faktorenanalyse an. Die Faktoren ergeben dann bestimmte ‚Ladungsmuster‘.

Von McQuitty (1954a) wurde eine Modifikation des Übereinstimmungsscores von Zubin (1938) vorgeschlagen. Im Gegensatz zu Zubin (1938) wird für jedes Paar von Individuen als Maßzahl nicht mehr die Anzahl der Items genommen, in der das Paar positiv übereinstimmt, sondern zu dieser Zahl wird noch die Zahl der Items addiert, in der das Paar ‚negativ‘ übereinstimmt. Falls für ein Item die Antworten 1 (ja'), 2 (weiß nicht') und 3 (nein') möglich sind, und auf das erste Item antworten beide Individuen eines Paares mit 1, während auf das zweite Item ein Individuum mit 1 und eines mit 3 antwortet, so berechnet sich der Übereinstimmungsscore auf die folgende Weise. Für das erste Item ergeben sich drei Übereinstimmungen, denn beide antworten mit 1, beide

antworten nicht mit 2 und beide antworten nicht mit 3. Für das zweite Item ergibt sich nur eine Übereinstimmung, weil nämlich beide nicht mit 2 antworten. Die Summe der Scores über alle Items, die hier gleich  $3 + 1 = 4$  ist, ergibt den Übereinstimmungsscore. Das Verfahren ist auch für den Fall von Items mit unterschiedlich vielen Alternativantworten anwendbar. Die Personen des oder der Paare mit dem höchsten Übereinstimmungsscore bilden die sogenannte erste *Spezies*. In der symmetrischen *Übereinstimmungsmatrix*, die analog zu der Ähnlichkeitstafel von Zubin (1938) aufgestellt wird, werden alle Zeilen und Spalten gestrichen, die sich auf diese Personen beziehen und in der Restmatrix die zweite *Spezies* bestimmt, die dem jetzt höchsten Übereinstimmungsscore entspricht. Dieses Verfahren wird immer weiter fortgesetzt. Falls alle Paare unterschiedliche Scores haben, umfassen die Spezies jeweils 2 Individuen, bis auf die letzte, die bei ungeradem Stichprobenumfang nur aus einem Individuum besteht.

Als nächstes wird für jede Spezies ihr Reaktionsmuster bestimmt. Dazu vereinigt man Antwortkategorien eines Items, in dem Individuen einer Spezies nicht übereinstimmen, zu einer neuen Kategorie, in der zwischen diesen Antworten nicht mehr Unterschieden wird. Auf diese Weise werden die Zahlen der Antwortkategorien reduziert und die Mitglieder einer Spezies stimmen in den Beurteilungen der verbleibenden Kategorien überein. Mit den so erhaltenen Reaktionsmustern der Spezies erstellt man wieder eine Überstimmungsmatrix und erhält durch Zusammenfassung von Spezies sogenannte ‚Genera‘. Aus diesen erhält man auf die gleiche Weise ‚Familien‘, daraus ‚Ordnungen‘, daraus ‚Klassen‘ usw., bis alle Individuen zusammengefaßt sind.

Das obige Verfahren verwendet McQuitty (1954a), um Kranke und Gesunde mit Hilfe von Antwortmustern zu klassifizieren. Dazu wird eine Krankenstichprobe und eine Gesundenstichprobe jeweils in eine Experimental- und eine Kreuzvalidierungsstichprobe aufgeteilt. Für jede Experimentalstichprobe werden die Individuen in Spezies, Genera usw. klassifiziert. Dieses ergibt zwei getrennte Klassifikationsschemata für Gesunde und Kranke. Jedes Individuum der Kreuzvalidierungsstichproben wird nun klassifiziert, indem man es dem Schema zuweist, in dem es die höchsten Übereinstimmungsscores hat.

In McQuitty (1954b) werden Methoden für eine objektive Einschätzung psychologischen Wohlbefindens gesucht. Dabei werden als typenanalytische Verfahren eine Faktorenanalyse für die Muster gemäß McQuitty (1953) und eine Übereinstimmungsanalyse gemäß McQuitty (1954a) diskutiert.

In McQuitty (1956) wird die Übereinstimmungsanalyse erstmals systematisch entwickelt. Ausgangspunkt ist das Meehl'sche Paradoxon nach Meehl (1950), der darauf hinwies, daß in der Kombination von Variablen bei der Vorhersage von Prädiktoren mehr Information enthalten sein kann als in den Einzelvariablen.

Für die verallgemeinerte Übereinstimmungsanalyse macht McQuitty (1956) folgende Grundannahmen:

1. Die Reaktionsmuster eines Individuums auf die Items eines Tests geben Aufschluß über m Kategorien von Individuen, wobei m durch die Analyse bestimmt wird. Die Kategorien bilden eine Hierarchie beginnend bei den individuellen Mustern, über Spezies, Genera, Familien, Ordnungen, Klassen usw., wobei jede Kategorie durch Zusammenfassung der früheren Kategorien entsteht.
2. Eine Reaktion auf ein Item ist entweder relevant oder nicht relevant für eine Kategorie. Eine Reaktion ist relevant genau dann, falls jedes Muster der Kategorie sie zeigt. Sonst ist sie irrelevant. Falls ein Item eine Reaktion aufweist, die für eine Kategorie relevant ist, so ist das Item selbst relevant für die Kategorie. Falls alle Reaktionen irrelevant sind, so ist das Item irrelevant. Das *Kategoriemuster* bilden die Reaktionen der relevanten Items, die in den Mustern enthalten sind, die in der Kategorie klassifiziert sind. Alle diese Muster haben dieselbe Reaktion für jedes relevante Item.
3. Reaktionen auf irrelevante Items sind zufällig.

Aus diesen Annahmen leitete McQuitty (1956) folgende *Grundgleichung der Übereinstimmungsanalyse* her:

$$n'_{ij} = n_{ij} - \frac{r - n_{ij}}{k^{p+q-1} - 1} .$$

Dabei ist  $n_{ij}$  der Übereinstimmungsscore für die Muster i und j gemäß Zubin (1938), d.h. die Anzahl der Items, für die Muster i dieselben Reaktionen wie Muster j zeigt. Ferner ist r die Gesamtzahl der Items im Test, p die Anzahl der individuellen Muster, die in die Kategorie mit dem Muster i klassifiziert wurden, q die Anzahl der individuellen Muster, die in die Kategorie mit dem Muster j klassifiziert wurden und k die für alle Items als gleich angenommene Anzahl der Antwortalternativen. Aus diesen Angaben erhält man den korrigierten Übereinstimmungsscore  $n'_{ij}$  für die Muster i und j, d.h.  $n_{ij}$  minus der Anzahl der irrelevanten Items, für die die Muster per Zufall übereinstimmen.

Um zu bestimmen, welche zwei individuellen Muster den höchsten korrigierten Übereinstimmungsscore haben, wird eine Matrix berechnet, die diesen Score für jedes Musterpaar enthält. Eine Vorform  $S_1$  der ersten Spezies besteht dann aus den individuellen Mustern mit den höchsten Scores. Als nächstes wird das Muster von  $S_1$  gemäß Annahme 2 bestimmt. Es besteht aus den Reaktionen der Items, für die die Muster übereinstimmen. Dieses Speziesmuster sei  $SP_1$ . Mit Hilfe der Grundgleichung wird der korrigierte Übereinstimmungsscore für jedes individuelle Muster mit  $SP_1$  bestimmt. Die ursprüngliche Übereinstimmungsmatrix wird um diese Scores durch Hinzunahme einer neu-

en Spalte und Zeile erweitert. Danach wird der größte Score in der erweiterten Matrix aufgesucht. Falls er sich für bisher nichtklassifizierte individuelle Muster ergibt, bilden diese eine Vorform  $S_2$  einer neuen Spezies, die zu einem Speziesmuster  $SP_2$  führt. Falls sich der Score zwischen  $SP_1$  und einem bisher nichtklassifizierten individuellen Muster ergibt, ist  $S_1$  erweitert und  $SP_1$  verfestigt worden.

Im Falle der Bildung von  $SP_2$  werden die Scores von  $SP_2$  mit jedem individuellen Muster und mit  $SP_1$  gebildet und die Matrix entsprechend erweitert. Danach wird wieder der größte Score aufgesucht. Falls sich dieser zwischen  $SP_2$  und einem individuellen Muster  $z$  ergibt, so wird  $S_2$  zu  $S_2^1$  vergrößert mit dem Speziesmuster  $SP_2^1$ . Die Veränderung des Speziesmuster von  $SP_2$  zu  $SP_2^1$  kann so groß sein, daß die ursprünglich zu  $S_2$  gehörenden Individuen nicht mehr durch  $SP_2^1$  optimal klassifiziert werden. Um dieses festzustellen, bestimmt man die Scores der individuellen Muster von  $SP_2^1$ . Falls eines dieser Muster seinen maximalen Score mit einem Muster außerhalb von  $S_2^1$  hat, so wird  $z$  wieder aus  $S_2^1$  entfernt, bis es, ohne eine Inkonsistenz zu verursachen, zu  $S_2$  hinzugenommen werden kann. Nachdem alle individuellen Muster in Spezies klassifiziert worden sind, werden diese nach der gleichen Methode zu Genera zusammengefaßt usw. Es werden auch Versuche unternommen, festzustellen, auf welchem Klassifikationsniveau die Übereinstimmungsscores die höchste statistische Signifikanz haben. Auf diesem Niveau sollen dann die realen Typen abgebildet werden. Man vergleiche hierzu McQuitty (1954c, 1957c).

In McQuitty (1956) wird darauf hingewiesen, daß diese Art von Übereinstimmungsanalyse analog zur Q-Faktorenanalyse im Sinne von Cattell (1952) konstruiert ist, daß aber Analoga zu den R-, P-, O-, S- und T-Techniken ebenfalls konstruierbar sind. Ferner werden Möglichkeiten diskutiert, das Verfahren abzukürzen, z.B. falls man nur Personen nach einem Außenkriterium klassifizieren will.

Einen Vergleich der Methode von McQuitty (1956) mit der Faktorenanalyse für qualitative Daten von Burt (1950) führt Watson (1956) anhand eines Datensatzes durch. Dabei zeigt sich, daß die Hauptkategorien in beiden Analysen einander entsprechen. Einen anderen Vergleich der beiden Verfahren findet man bei McQuitty (1957c).

In McQuitty (1957a) wird ein Musteranalyseverfahren für ungeordnete Daten vorgeschlagen für den Zweck, die Validität von Tests für Prädiktionskriterien zu erhöhen. Eine Einordnung des Verfahrens erfolgt nach der Einteilung von Tiedemann (1955), der zwischen Musteranalyseverfahren für geordnete Daten, z.B. der Profilanalyse nach Cronbach & Gleser (1953), und solchen für ungeordnete Daten unterscheidet. Die Musteranalyseverfahren für ungeordnete Daten lassen sich unterteilen in *kumulative Verfahren*, *reduktive Verfahren* und *Dualmusterverfahren*. Im kumulativen Ansatz werden Reaktionsmuster

seriell aufgebaut. In McQuitty (1957a) werden die Vorteile und Nachteile dieses Ansatzes, den man bei Lubin & Osburn (1957) beschrieben findet, dargestellt. Im Gegensatz zu den kumulativen Reaktionsmustern, die seriell aufgebaut werden, indem ein Item nach dem anderen hinzugenommen wird, erhält man reduktive Reaktionsmuster, indem man mit allen Reaktionen eines Individuums auf die Items eines Tests beginnt und dieses individuelle Reaktionsmuster auf ein oder mehrere Muster mit verringrigerer Itemzahl reduziert. Dieses geschieht in der Übereinstimmungsanalyse von McQuitty (1956) und in der *Multi-Prototyp-Analyse* von McQuitty (1955a, 1958). In dem letzteren Verfahren werden die Hauptreaktionsmuster isoliert, die für jede Kriteriumsgruppe von Individuen eindeutig bestimmt sind.

Neben kumulativen und reduktiven Mustern betrachtet McQuitty (1957a) Dualmuster. Hierzu werden zunächst Individuen aufgrund der Scores von verschiedenen Kriterien in Kategorien klassifiziert. Für jede Kriterienkategorie von Individuen wird dann ein zugehöriger Test durchgeführt, der ein duales Testreaktionsmuster ergibt. Diese Muster werden als Prädiktoren der Kriterienmuster für andere Stichproben verwendet.

Ein weiteres Typenanalyseverfahren ist die *elementare Verbindungsanalyse* von McQuitty (1957b). Der Hauptvorteil gegenüber den bei Cattell (1944) beschriebenen Clusterverfahren wird darin gesehen, daß die Hinzunahme eines Individuums zu einem Cluster nicht nach einem willkürlichen Distanzkriterium erfolgt, sondern durch die Assoziation des Individuums mit dem Cluster, die durch ein datenabhängiges nicht willkürliche Kriterium bestimmt wird. Ausgangspunkt ist eine Typentheorie, die von einer Typenstruktur ausgeht, in der jedes Mitglied eines Typs einem andern Mitglied desselben Typs ähnlicher ist als einem beliebigen Mitglied eines anderen Typs. Nach der Isolierung der Mitglieder eines Typs definiert man einen *Prototyp* als die Gesamtheit der Charakteristika, die die Mitglieder des Typs besitzen. Als *Relevanz* des Mitgliedes eines Typs bezüglich des zugehörigen Prototyps wird die zugehörige Ähnlichkeit bezeichnet.

Als Ähnlichkeitsmaß werden Korrelationskoeffizienten verwendet. Jede Person eines Typs hat eine höhere Korrelation mit irgendeiner anderen Person des Typs als mit jemandem außerhalb des Typs. Es seien i und j zwei Individuen, so daß i mit j die höchste Korrelation hat und j mit i. Dann heißen i und j ein *reziprokes Paar*. Solche Paare gehören nach Definition zu demselben Typ. Man kann zeigen, daß es immer mindestens ein reziprokes Paar gibt. Nachdem man alle reziproken Paare, die als Kernelemente der Typen angesehen werden, identifiziert hat, ordnet man sukzessiv alle noch nicht verwendeten Individuen denjenigen Typen zu, in denen sich jeweils das Individuum befindet, mit dem sie die höchste Korrelation haben.

Es werden in McQuitty (1957b) mehrere Techniken beschrieben, um für die verschiedenen Typen *orthogonale* Kerne zu erhalten. Eine andere Erweiterung

ist die *Centroid-Verbindungsanalyse*. Hier wird jeweils die maximale Ähnlichkeit mit einem Kernelement zum Anlaß genommen, ein Individuum zu einem Typ hinzuzurechnen. In der *hierarchischen Verbindungsanalyse* werden die Typen, jetzt als Spezies bezeichnet, zum Ausgangspunkt einer neuen Verbindungsanalyse genommen, um sogenannte Genera zu erhalten. Das Verfahren kann man dann analog zur Übereinstimmungsanalyse in McQuitty (1956) fortsetzen. Auf dieses Verfahren wird noch eingegangen. In der *differentiellen Verbindungsanalyse* werden die Centroid- oder die elementare Verbindungsanalyse verwendet, um Reaktionsmuster auszuwählen, die zwischen zwei oder mehr Kategorien von Individuen differenzieren. Dazu führt man für jede Kategorie, etwa für Gesunde und Kranke, eine Verbindungsanalyse durch und bestimmt die für die einzelnen Typen charakteristischen Muster. Ein anderes Verfahren, das von der elementaren Verbindungsanalyse ausgeht, ist die sogenannte elementare *Faktorenanalyse* von McQuitty (1961a). Die Vorteile gegenüber der üblichen Faktorenanalyse werden in der Möglichkeit einer schnellen und elementaren Rechnung auch bei vielen Variablen gesehen.

In Forehand & McQuitty (1959) wird für eine Reihe von Prädiktoren eine Hauptkomponentenanalyse durchgeführt, von der die ersten vier Faktoren betrachtet werden. Dann werden die stetigen Scores für jeden Faktor trichotomisiert und jedes Individuum erhält auf diese Weise einen Vierervektor mit Komponenten, die die Werte 1, 2 und 3 annehmen können, zugewiesen. Danach werden 3 Verfahren miteinander verglichen, um aufgrund der Muster eine Kriteriumsvariable vorherzusagen. Das erste Verfahren verwendet einen multiplen Regressionsansatz. Das zweite Verfahren, als *Konfigurationsskalenprädiktion* bezeichnet, beruht auf den Ansätzen von Lubin & Osburn (1957) und Lykken (1956) und geht von den Gruppen von Individuen mit identischen Mustern aus. Das dritte Verfahren schließlich, als *Signifikanzmustervorhersage* bezeichnet, betrachtet nicht nur die Muster für vier Faktoren, sondern auch die Muster für alle Kombinationen von drei, zwei und einem Faktor. Es werden nur die Muster betrachtet, die unter der Annahme der totalen Unabhängigkeit und der Gleichwahrscheinlichkeit aller Alternativen eine bei einer Alpha-Schranke von 10% zu hohe Häufigkeit aufweisen. Diese *Schlüsselmuster* werden verwendet, um die Individuen zu klassifizieren. Der vorhergesagte Kriteriumswert für ein Individuum ergibt sich als Mittelwert der Kriteriumswerte für alle Individuen mit dem gleichen Schlüsselmuster. Im empirischen Vergleich zeigte zwar die Konfigurationsskalenprädiktion die größte Validität, jedoch war bei einer Kreuzvalidierung der multiple Regressionsansatz überlegen. Ursachen werden in zu kleinen Stichprobenumfängen vermutet.

Die schon in McQuitty (1957 b) angedeutete Möglichkeit einer *hierarchischen Verbindungsanalyse* wird in McQuitty (1960a) weiter ausgeführt. Es handelt sich um eine Verknüpfung der Ideen der elementaren Verbindungsanalyse und der Übereinstimmungsanalyse. Ausgangspunkt ist eine Interassoziationsma-

trix zwischen Individuen, z.B. eine Interkorrelationsmatrix oder eine Übereinstimmungsscorematrix. Grundlegend für die elementare Verbindungsanalyse ist der Begriff des reziproken Paares (i,j), für welches der Assoziationsindex unter allen Paaren von Individuen, in denen i oder j enthalten ist, der höchste ist. In der hierarchischen Verbindungsanalyse verwendet man analog das Konzept der *reziproken Kombinationen*. Dabei bilden n Individuen eine reziproke Kombination, falls jedes Individuum der Kombination mindestens soviel mit k anderen Individuen der Kombination gemeinsam hat wie mit k Individuen, die nicht alle zur Kombination gehören, wobei k die Werte von 1 bis (n-1) durchläuft. Reziproke Kombinationen sind genau dann nicht überlappend, falls jedes Individuum irgendeiner Kombination zu keiner Kombination höherer Ordnung mit mehr Individuen gehört, falls die anderen Individuen seiner Kombination nicht ebenfalls dazu gehören.

Der Zweck der hierarchischen Verbindungsanalyse besteht darin, eine Gruppe von Individuen so in hierarchische Kategorien zu klassifizieren, daß die Mitglieder jeder Kategorie eine maximale Zahl gemeinsamer Charakteristika haben und gleichzeitig die Zahl der benötigten Kategorien minimal ist. Die Klassifikation faßt schrittweise Individuen bzw. Gruppen von Individuen zusammen. Dabei werden nur die besten der möglichen Klassifikationen tatsächlich beibehalten. Je nachdem, ob nur nichtüberlappende Kategorien oder auch überlappende Kategorien zugelassen sind, unterscheidet McQuitty (1960 a) zwischen einer umfassenden und einer kurzen Version des Verfahrens.

Eine weitere Methode zur Klassifikation von Personen zu Typen ist die *hierarchische Syndromanalyse* von McQuitty (1960 b). Ausgangspunkt ist wieder eine Interassoziationsmatrix, gebildet etwa aus den Übereinstimmungsscores von Zubin (1938). Für eine bestimmte Kategorie wird angenommen, daß alle Individuen der Kategorie in sovielen Charakteristika übereinstimmen wie dasjenige Paar von Individuen, das die wenigsten Übereinstimmungen aufweist (*Klassifikationsannahme*).

In der sogenannten *Ersetzungsversion* werden vier Iterationsschritte so oft angewandt, wie es Spalten in der ursprünglichen Matrix gibt. Die erste Anwendung geschieht bei der Originalmatrix und die folgenden bei Matrizen, die dadurch entstehen, daß jeweils 2 Spalten mit den zugehörigen Zeilen in der jeweils vorangehenden Matrix gestrichen werden, aus denen dann eine neue Spalte bzw. Zeile gebildet wird. Im ersten Schritt sucht man das höchste Element der Matrix, das die Assoziation zwischen den Individuen i und j angeben möge. Im zweiten Schritt werden i und j vereinigt zu einer Kombination ij, für die man eine neue Zeile und Spalte vorsieht. Im dritten Schritt wählt man den kleineren der beiden Werte, die zu (Zeile k, Spalte i) bzw. (Zeile k, Spalte j) gehören und trägt ihn ein bei (Zeile k, Spalte ij) und (Zeile ij, Spalte k). Im vierten Schritt streicht man die Zeilen und Spalten für die Individuen i und j.

Falls in der Ersetzungsversion zwei Individuen i und j einmal zusammengefaßt wurden, so bleiben sie auch in allen höheren Stufen der Klassifikation zusammen, ohne zu berücksichtigen, daß die frühere Klassifikation möglicherweise auf nicht ausreichende Information zurückzuführen ist. Eine vollständigere Version, die sogenannte *selbstkontrollierende* Version, berücksichtigt dieses, indem immer nur die Spalten, aber nicht die Zeilen eliminiert werden. Falls die Daten durch nichtüberlappende Typen erklärt werden, stimmen beide Versionen überein.

Eine Weiterentwicklung des Verfahrens ist die *umfassende hierarchische Analyse* von McQuitty (1960 c). Hierbei werden Individuen so in Kategorien klassifiziert, daß es für jede Kombination von Individuen eine eigene Kategorie gibt und zu dieser Kategorie einen Index der Zahl der gemeinsamen Charakteristika. Neben der Version, die auf einer Interassoziationsmatrix beruht, wird auch noch eine sogenannte Rangordnungsversion betrachtet. Hierbei werden alle Assoziationskoeffizienten für Paare in eine Rangordnung gebracht und allen möglichen Kombinationen gegenübergestellt. Das Paar mit dem Rang 1 (niedrigster Assoziationskoeffizient) wird ausgewählt und alle Kombinationen, in denen es auftritt, werden unter dieses Paar geschrieben. Dann führt man dasselbe durch für das Paar mit dem Rang 2, wobei man die verwendeten Kombinationen nicht noch einmal benutzen darf, usw. Diese Rangordnungsversion liefert dasselbe Ergebnis wie die Interassoziationsmatrixversion.

Um den Rechenaufwand herabzusetzen, sieht McQuitty (1960 c) folgende Möglichkeiten. Einmal kann man sich darauf beschränken, nur die typverdächtigen Kombinationen zu betrachten. Eine weitere Reduktion ergibt sich dadurch, daß man Kombinationen mit maximalen Übereinstimmungsscores zu demselben Typ zusammenfaßt, wobei es mehrere Möglichkeiten gibt, die maximale Übereinstimmung von Kombinationen zu definieren.

In McQuitty (1957a, 1959) werden Methoden angegeben, um Individuen aufgrund ihrer Antwortkonfigurationen auf die Items eines Tests zu klassifizieren. Hierbei werden die vorherrschenden Muster, die die Anzahl der Items maximieren, durch die die Personen klassifiziert werden, isoliert. Dieses hat für diagnostische Zwecke den Nachteil, daß einige oder alle möglichen Diagnosen mit nicht vorherrschenden Mustern verknüpft sein können. Es scheint dann besser zu sein, diejenigen Muster zu isolieren, die z.B. für Patienten bzw. Gesunde typisch sind, unabhängig davon, ob sie vorherrschend sind oder nicht. Eine solche Methode schlägt McQuitty (1961 b) vor. Dazu wird für jede Diagnosegruppe eine hierarchische Syndromanalyse oder eine umfassende hierarchische Analyse durchgeführt und festgestellt, welche Kategorien nur bei jeweils einer Kriteriumskategorie (Diagnose) auftreten und die dadurch gekennzeichneten *differentiellen Muster* für diagnostische Zwecke verwendet.

Ein Verfahren, um Items zu diagnostischen Zwecken auszuwählen, gibt McQuitty (1961 d) an. Dazu betrachte man eine Gruppe von Kranken und eine

Gruppe von Gesunden und führe für jede Gruppe getrennt eine Klassifikation in Kategorien nach irgendeiner musteranalytischen Methode, z.B. nach McQuitty (1957b, 1959, 1960a, 1960b, 1960c, 1961 c), durch. Jede Kategorie der Gesunden wird mit jeder Kategorie der Kranken gepaart. Alle Items, die innerhalb eines Kategorienpaares gut differenzieren, bilden einen Test für dieses Paar. Soll die Gruppe der Kranken einen höheren Score liefern, so erhält für jeden der Tests die Antwortalternative, die den Kranken entspricht, einen Score 1 und die andere Alternative den Score 0. In einer Kreuzvalidierungsstichprobe wird die Validität der Tests überprüft.

In McQuitty (1961 c) wird ein Typ definiert als eine Gruppe von n Personen, bei der jede Person jeder der (n-l) anderen Personen des Typs ähnlicher ist als einer Person außerhalb des Typs. Ausgehend von einer Interassoziationsmatrix gibt McQuitty (1961 c) einen Algorithmus an, mit dem er derartige Typen isolieren kann. Dazu wird jede Assoziation  $r_{ij}$  für jedes x mit  $r_{ix}$  und  $r_{jx}$  verglichen. Als Vorteil dieser sogenannten *Typenanalyse* wird angegeben, daß sie dort Typen innerhalb von Typen auffinden kann, wo dies die inverse Faktorenanalyse und die elementare Verbindungsanalyse nicht können.

Eine Erweiterung der hierarchischen Syndromanalyse von McQuitty (1960 b) stellt die sogenannte *multiple hierarchische* Syndromanalyse von McQuitty (1962) dar. In dieser erweiterten Form werden Individuen nicht nur gemäß ihren vorherrschenden Mustern klassifiziert, sondern es werden zusätzlich noch die nächstfolgenden vorherrschenden Muster mit herangezogen. Dazu werden allerdings nur Antwortmuster verwendet, die in der ersten Klassifikation nicht verwendet wurden. Mit diesem Verfahren fahrt man fort, bis jedes Individuum bezüglich aller seiner Antworten klassifiziert ist. Man unterscheidet auch hier wieder eine *Ersetzungs-* und eine *selbstkontrollierende* Version. Die Ersetzungsversion wird für die vorherrschenden Muster verwandt, die selbstkontrollierende Version für die folgenden Muster. Die Klassifikation verwendet drei Prinzipien: 1. Die *Klassifikationsannahme*, d. h. jede Kategorie hat soviel gemeinsam wie das Paar, das am wenigsten gemeinsam hat. 2. Das Prinzip der *Maximumklassifikation*, welches besagt, daß diejenigen Kategorien als signifikante Kategorien verwendet werden, für die das Produkt aus Übereinstimmungsscore und Personenzahl maximal ist. 3. Die Berechnung von *Übereinstimmungsscoreresiduen*, die man erhält, wenn man von dem Übereinstimmungsscore für ein Paar ij den höchsten Übereinstimmungsscore einer signifikanten Kategorie abzieht, in die i und j klassifiziert werden.

Eine Erweiterung der Rangordnungsversion von McQuitty (1960 c) in Verbindung mit der Typenanalyse von McQuitty (1961 c) stellt die *Rangordnungstypenanalyse* von McQuitty (1963 a) dar. Bei dieser werden in der Interassoziationsmatrix in jeder Spalte die Assoziationskoeffizienten durch ihre Ränge ersetzt. Eine Untermatrix dieser Matrix, die genau aus den zu einer Teilmenge von Individuen gehörenden Rängen besteht, stellt dann einen Typ nach der

Definition von McQuitty (1961 c) dar, falls sie keinen Rang enthält, der größer als die Anzahl der Individuen der Teilmenge ist. Es wird eine Prozedur angegeben, um alle diese Untermatrizen anzugeben.

Sowohl die hierarchische Syndromanalyse von McQuitty (1960 b) als auch die multiple hierarchische Analyse von McQuitty (1962) sorgen nicht unbedingt dafür, daß ein Individuum auf jeder Stufe optimal klassifiziert wird. Sie klassifizieren zunächst ein Individuum optimal in eine Spezies und klassifizieren dann die Spezies optimal in einen Genus. Deshalb schlägt McQuitty (1963 b) eine Methode vor, die jedes Individuum zunächst optimal in eine Gruppe von zwei, dann drei usw. Personen klassifiziert.

In McQuitty (1960 a) wird die hierarchische Verbindungsanalyse als eine hierarchische Version der elementaren Verbindungsanalyse von McQuitty (1957 b) vorgeschlagen. Andere hierarchische Versionen der Verbindungsanalyse werden in McQuitty (1964) diskutiert. Eine Möglichkeit besteht darin, nach Durchführung einer elementaren Verbindungsanalyse für jede Kategorie ein *Referenzitem* zu bestimmen und mit diesen Items das Verfahren fortzusetzen. Eine Verfeinerung sieht die Hinzunahme weiterer Items vor. Schließlich wird auch die Hinzunahme aller Items erwogen. Um auch auf höheren Stufen des Klassifikationsprozesses Assoziationsindizes zur Verfügung zu haben, verwendet McQuitty (1964) die Klassifikationsannahme, den Ähnlichkeitsindex von McQuitty (1955 b) und den korrigierten Übereinstimmungsscore von McQuitty (1956).

Wenn man davon ausgeht, daß Individuen verschiedenartigen Einflüssen ausgesetzt sind, liegt es nahe, verschiedene einander überlagernde Typologien anzunehmen. Um derartige unterschiedliche Typologien trennen zu können, schlägt McQuitty (1965) eine *multiple Rangordnungstypenanalyse* vor. Hierzu wird eine Rangordnungstypenanalyse (McQuitty, 1963 a) durchgeführt und für jeden Typ die Itemmenge bestimmt, in der die Individuen dieses Typs übereinstimmen. Für jede dieser Itemmengen wird für die Gesamtmenge der Individuen eine Interassoziationsmatrix berechnet und auf diese eine Rangordnungstypenanalyse angewandt. Auf diese Weise erhält man neue Typen für jeden der ursprünglichen Typen.

Eine Überlagerung von verschiedenen Typen muß nicht notwendig zu *Überschneidungstypen* führen, wie in McQuitty (1965) angenommen wird. Falls die Einflüsse auf ein Individuum unabhängig voneinander sind, ist es möglich, daß sich unabhängige Typen und Typologien für diesen Fall herausbilden. In McQuitty (1966 a) werden Versionen der multiplen Rangordnungstypenanalyse zur Isolation unabhängiger Typen beschrieben. Zunächst wird dazu wieder eine Rangordnungstypenanalyse auf eine Interassoziationsmatrix angewandt. In der ersten Version werden dann *maximalumfassende Typen* definiert. Dieses ist die Minimalzahl von benötigten Typen, um alle Individuen zu umfas-

sen, die in diesem Stadium zu Typen gehören. Für jeden dieser Typen werden dann die gemeinsamen Charakteristika für die Beschreibung der Individuen dieses Typs bestimmt und herausgenommen. Für die Restitems wird eine neue Interassoziationsmatrix berechnet und darüber wieder eine Rangordnungstypenanalyse durchgeführt, usw. Eine andere Version verwendet anstelle der maximalumfassenden Typen die Typen erster *Ordnung*, die sich bei der ersten Anordnung der Rangordnungstypenanalyse ergeben.

In McQuitty (1966 b) wird zwischen realen *unvollkommenen Typen* und theoretischen *reinen Typen* unterschieden, wobei die reinen Typen aus zwei oder mehr unvollkommenen Typen zusammengesetzt sind. Als Zwischenstufe zwischen reinen und unvollkommenen Typen sind *hierarchische Typen* anzusehen, die sich bei der Klassifikation von unvollkommenen Typen in immer größere in sich konsistente Kategorien ergeben. Vor diesem Hintergrund wird die *hierarchische Analyse durch reziproke Paare* eingeführt. Bei dieser wird zunächst eine Matrix der Assoziationsindizes zwischen allen unvollkommenen Typen aufgestellt und die reziproken Paare isoliert. Für jedes reziproke Paar werden die gemeinsamen Charakteristika spezifiziert und dadurch hierarchische Typen definiert, die anstelle der sie definierenden unvollkommenen Typen eingesetzt werden. Es wird eine neue Matrix für die Assoziationen zwischen den hierarchischen Typen und den nichtverwendeten unvollkommenen Typen berechnet und das Verfahren iterativ wiederholt. Um unabhängige Typen zu isolieren, wird die sogenannte *multiple hierarchische Analyse durch reziproke Paare* vorgeschlagen (McQuitty, 1966 b), in der zusätzlich das Prinzip der Maximum-Klassifikation (McQuitty, 1955c, 1957a, 1962) verwendet wird. Ohne Verwendung dieses letzteren Prinzips erhält man eine *multiple hierarchische Analyse durch reziproke Paare für die Isolation von Überschneidungstypen*.

Die hierarchische Syndromanalyse (McQuitty, 1960b) hat den Nachteil, daß auf einer Stufe der Klassifikation zusammengefaßte Individuen auf einer höheren Stufe nicht mehr getrennt werden können. Die bisherigen Verbesserungsvorschläge führten zu relativ komplizierten Klassifikationen. Deshalb schlägt McQuitty (1966 c) eine Verbesserung vor, die die Typentheorie aus McQuitty (1966 b) bei einer neuen Version der hierarchischen Syndromanalyse aus McQuitty (1960b) verwendet. Diese sogenannte *Höchstes-Paar-Version* der hierarchischen Syndromanalyse ist einfacher als die Ersetzungsversion und die selbstkontrollierende Version, wobei alle drei Versionen auf diskrete und auf stetige Daten anwendbar sind. Die Grundidee ist, jeweils ein Paar von unvollkommenen Typen zu einem hierarchischen Typ zusammenzufassen, wobei man vom größten Assoziationskoeffizienten in der Interassoziationsmatrix ausgeht und die Matrix so schrittweise verkleinert.

In McQuitty (1955 b) wird ein Ähnlichkeitsindex für die sogenannte Ähnlichkeitsanalyse entwickelt. Da die ursprüngliche Form dieser Analyse kompli-

ziert ist und zu Widersprüchen führen kann, schlägt McQuitty (1966d) die sogenannte *Ähnlichkeitsanalyse durch reziproke Paare* vor, die auf diskrete und stetige Daten anwendbar ist und auf einer Verwendung der Typentheorie von McQuitty (1966 b) beruht. Der Assoziationsindex zwischen einer Typenrepräsentantenkombination ij und einem Typenrepräsentanten k, der von i und j verschieden ist, wird dabei als Mittel der Assoziationsindizes zwischen i und k sowie j und k berechnet. In McQuitty (1967b) werden zwei verschiedene Möglichkeiten diskutiert, die Individuen bei der Berechnung der Assoziationsindizes zu gewichten. Das Verfahren liefert hierarchische Typen verschiedener Ordnung.

In einem Übersichtsartikel von McQuitty (1967a) werden die Typentheorie, die Verbindungsanalyse, die Rangordnungstypenanalyse, die hierarchische Klassifikation durch reziproke Paare, die hierarchische Klassifikation durch Rangordnungstypen und die Isolation von Überschneidungstypen und unabhängigen Typen aufeinander bezogen und der logische Entwicklungsprozeß dargestellt. Anwendungen und Illustrationen der Verfahren findet man in McQuitty (1967d) und Schubert (1965). In McQuitty (1967c) wird empfohlen, neben den typentheoretischen Ansätzen auch die Isolierung von dimensionalen Konstrukten parallel durchzuführen, wobei für beide Ansätze eine auf Produktmomentkorrelationskoeffizienten beruhende Methode vorgeschlagen wird.

In der hierarchischen Klassifikation durch reziproke Paare (McQuitty, 1966 b) und ähnlichen Verfahren tritt gelegentlich das Problem von *Bindungen* auf, wenn Assoziationsindizes den gleichen Wert haben. Für diesen Fall geben McQuitty, Price & Clark (1967) eine allgemeine Lösung an. Falls etwa die Assoziationen für ij und ik gleich und maximal sind, während die für jk relativ klein ist, so kann es sein, daß der Typ ijk auf keiner Stufe realisiert wird. Der Vorschlag ist, sowohl den Typ ij als auch den Typ ik zu bilden.

Eine Schwäche der bisherigen Ansätze der Musteranalyse wird von McQuitty (1968 a) darin gesehen, daß an der Entscheidung, wann eine neue Kategorie gebildet werden soll, immer nur relativ wenige Assoziationsindizes beteiligt sind. Es wird deshalb vorgeschlagen, die Elemente der Interassoziationsmatrix durch ihre Ränge zu ersetzen und die Assoziation zwischen i und j durch den Rangkorrelationskoeffizienten von Spearman für die Spalten i und j zu messen, gegebenenfalls mit Bindungskorrektur. Da zu den Elementen (i,j) bzw. (j,i) jeweils das Element (i,i) bzw. (j,j) fehlt, wird das zu (i,j) und (j,i) gehörige Rangpaar betrachtet. Diese Interspaltenkorrelationen werden sowohl mit der verbesserten hierarchischen Syndromanalyse für diskrete und stetige Daten von McQuitty (1966c) als auch mit der Rangordnungstypenanalyse von McQuitty (1963 a) analysiert.

Man ersetzt bei dieser Methode die ursprüngliche Interassoziationsmatrix durch eine Matrix von Rangkorrelationskoeffizienten. Falls man das Verfahren

auf diese Matrix iteriert anwendet, so ergibt sich die von McQuitty & Clark (1968a) beschriebene *iterative Interspaltenkorrelationsanalyse*. Es wird bewiesen, daß durch dieses Verfahren tatsächlich vorhandene Typen isoliert werden, d.h. daß die Korrelationskoeffizienten für Personen des gleichen Typs gegen 1 streben. Gegen -1 streben sie für Personen, die zu unterschiedlichen Typen gehören. Für nicht eindeutig zugeordnete Personen ergeben sich Zwischenwerte. Die Beweise lassen sich auch übertragen auf den Fall, daß eine Fehlervarianz in den Antworten zugelassen wird. Eine Anwendung des Verfahrens bei der Überprüfung der Reliabilität von intraindividuellen Persönlichkeitsstrukturen im Zusammenhang mit einem Test über psychisches Wohlbefinden geben McQuitty, Abeles & Clark (1970).

Eine Erweiterung der iterativen Interspaltenkorrelationsanalyse zu einer *multiplen iterativen Interspaltenkorrelationsanalyse* wird in McQuitty (1968 b) vorgenommen. Dieses Verfahren gestattet es, jede Person nach fast allen ihren Charakteristika zu klassifizieren und ist eine Adaption der hierarchischen Analyse durch reziproke Paare von McQuitty (1966 b).

Verschiedene Formen von Typen, die bei einer iterativen Interspaltenkorrelationsanalyse auftreten können, diskutieren Clark & McQuitty (1970). Ein Typ, der durch eine eindeutige Menge von Charakteristika definiert ist, die allen Individuen des Typs gemeinsam sind, heißt *Gruppentyp*. Ein Typ, der durch ein einzelnes Individuum als eindeutig verschieden zu anderen Einzelpersonen oder Gruppentypen bestimmt wird, heißt *Einzel-Objekt-Typ*. Ein Typ, der durch drei oder mehr Individuen definiert ist, die zwar keine Charakteristika aufweisen, die alle gemeinsam haben, aber einige Charakteristika, die einige Gruppenmitglieder gemeinsam haben, heißt *Gruppen-Quasi-Typ*. Schließlich bildet eine Gruppe von Individuen, die sich von allen anderen Individuengruppen unterscheidet und deren Charakteristika weder allen Individuen der Gruppe noch einigen Individuen der Gruppe gemeinsam sind, ein *System*. In Clark & McQuitty (1970) wird untersucht, welche Formen von Limesmatrizen zu welchen Typenformen führen.

Die Wahl eines Übereinstimmungsscores zwischen mehreren Personen kann davon abhängen, wie negative Antworten auf Items eines Tests gewichtet werden. Ein neuer Index für den Fall, daß negative Antworten unberücksichtigt bleiben, wird von McQuitty & Clark (1968 b) vorgeschlagen. Sie wählen den relativen Anteil der Charakteristika, der etwa den Personen 1, . . . , n gemeinsam ist. Dieser wird berechnet, indem man die Anzahl der gemeinsamen Charakteristika für alle n Personen bestimmt, diesen Wert mit n multipliziert und durch die Summe der Anzahlen der Charakteristika für die Einzelpersonen dividiert. Mit Hilfe einer hierarchischen Klassifikation durch reziproke Paare werden basierend auf diesem neuen Index in einem Anwendungsbeispiel sowohl Kurse klassifiziert gemäß den Studenten, die sie wählen, als auch Studenten klassifiziert gemäß den gewählten Kursen.

Ein Verfahren, das nicht nur den größten Assoziationsindex in einer Spalte der Interassoziationsmatrix berücksichtigt, sondern zusätzlich eine Minimalzahl der nächsthöheren Indizes nach Bedarf während des Klassifikationsprozesses mit heranzieht, ist die *hierarchische Klassifikation durch multiple Verbindungen* von McQuitty (1970a). Hierbei beruht jede Klassifikation auf einer oder mehreren Verbindungen mit dem zentralen Kern eines Clusters. Es gibt keine Zuordnungen, die nur auf Ketten von Individuen beruhen, die sukzessiv miteinander assoziiert sind. Das Verfahren kann sowohl Kategorien von unten her aufbauen als auch von oben her zerlegen. Es geht von einer Rangtransformation der Spalten der Interassoziationsmatrix aus und verwendet primär die höchsten Verbindungen zwischen den Individuen. Multiple Verbindungen werden nur bei Bedarf herangezogen.

Möglichkeiten, das Konzept der Reziprozität (McQuitty, 1967a) abzuschwächen, diskutiert McQuitty (1970 b). Es wird berücksichtigt, daß ein reziprokes Paar auch durch Zufall entstanden sein kann. Ein Paar von Individuen i und j heiße *reziprok*, falls i die höchste, zweithöchste usw. Assoziation mit j hat und j seinerseits die höchste, zweithöchste usw. Assoziation mit i hat. Auf diese Weise ergeben sich unterschiedliche Stufen der Reziprozität. Auf jeder Stufe der Analyse besteht damit die Möglichkeit, die Reziprozitätsstufe so weit herabzusetzen, daß reziproke Paare zugelassen werden, die sonst möglicherweise ausgeschlossen worden wären. Viele Fehlklassifikationen werden bei dieser *hierarchischen Klassifikation durch Mehrstufenreziprozität* auf einer höheren Stufe korrigiert. Falls i die k-höchste Assoziation mit j hat und j die l-höchste Assoziation mit i, so ist das Maximum von k und 1 die *Reziprozitätsstufe* von ij. Bei der Zuordnung der Ränge in jeder Spalte der Interassoziationsmatrix wird bei diesem Verfahren im Falle von Bindungen nicht die Mittelrangmethode benutzt sondern ein anderes Verfahren. Es gibt zwei Versionen des Verfahrens. Die Version der *sukzessiven Verbindungen* baut sukzessiv nach Bedarf die Reziprozitätsstufen ab. Die Version der *Kernzuweisungen* geht von in sich konsistenten Untermatrizen maximaler Ordnung der Rangmatrix für die Assoziationen aus. Diese Untermatrizen heißen auch Kerne.

Bisher (McQuitty, 1967a) war ein reiner Typ definiert als eine Kategorie von Individuen derart, daß jedes Individuum innerhalb der Kategorie zu jedem Individuum der Kategorie ähnlicher ist als zu einem beliebigem Individuum außerhalb. Alternativ kann ein *reiner Typ* nach McQuitty (1970c) definiert werden als eine Kategorie von Individuen mit einer eindeutigen Kombination von Charakteristika. Jedes Individuum in der Kategorie besitzt alle diese Charakteristika, während Individuen außerhalb der Kategorie nicht alle Charakteristika besitzen. Um solche reinen Typen zu identifizieren, wird eine hierarchische Klassifikation durch eine *Benennungs-Auswahl-Analyse* vorgenommen. Ausgangspunkt ist wieder eine Interassoziationsmatrix, deren Spalten rangtransformiert werden. Als erstes wird diese Matrix auf alle möglichen

Arten in zwei Untermatrizen zerlegt. Dann wird jedes Paar von Untermatrizen für alle möglichen Kombinationen von Individuen gebildet. Schließlich wird die spezielle Zerlegung und Kombination ausgewählt, die die Summe der durchschnittlichen Abweichungen innerhalb der Untermatrizen bezüglich ihrer zentralen Tendenzen minimiert. Wegen des großen Rechenaufwandes wird das Verfahren dadurch abgekürzt, daß jedes Individuum für jede Untermatrix eine Kombination von möglichst ähnlichen Individuen *benennt*, d.h. festlegt. Die Benennung, die das obige Kriterium bezüglich der Abweichungen von den zentralen Tendenzen minimiert, wird *ausgewählt*. Falls die Daten einen reinen Typ enthalten, wird er isoliert und angezeigt. Ansonsten wird angezeigt, daß eine perfekte Lösung nicht existiert und wie groß der Abstand zu ihr ist.

In McQuitty, Banks & Frary (1970) wird eine Methode beschrieben, um eine Interassoziationsmatrix mit rangtransformierten Spalten so zu unterteilen, daß man eine Untermatrix erhält, die dazu verwendet werden kann, den Grad der Abhängigkeit in der Ursprungsmatrix zu messen. Im Anschluß daran kann man dann die Bedeutsamkeit des Unterschiedes im Grad der Abhängigkeit bei zwei oder mehr Ursprungsmatrizen untersuchen.

Dieses Verfahren wird von McQuitty, Banks & Frary (1971) dazu verwendet, um Unterschiede zwischen gestörten und normalen Studenten aufzudecken. Es ergeben sich geringere Interassoziationen für die Gestörten im Vergleich zu den Normalen. Von McQuitty, Banks, Frary & Aye (1972) wird die Interassoziationsmatrix für zwei gleichgroße Stichproben von Gestörten und Normalen in alle möglichen Paare von Teilmatrizen gleicher Ordnung zerlegt. Es wird ein Kriterium angegeben, um ein solches Paar von Teilmatrizen auszuwählen, das in der einen Matrix vorzugsweise Gestörte und in der anderen Matrix Normale enthält. Aus den größten Übereinstimmungsscores werden dazu nach einer speziellen Vorschrift Mittelwerte für die Teilmatrizen gebildet und die Differenz dieser Mittelwerte als Abstand der Teilmatrizen verwendet. Mit Hilfe der Anzahlen der gestörten und nicht gestörten Individuen in jeder Teilmatrix kann die Bedeutsamkeit des Ergebnisses unter Verwendung der hypergeometrischen Verteilung beurteilt werden. Dieses Verfahren ist jedoch so rechenaufwendig, daß es nur für kleinere Stichproben anwendbar ist. Deswegen schlägt McQuitty (1971 c) eine abgekürzte Version vor. Bei dieser werden iterativ diejenigen Individuen eliminiert, die aufgrund sehr großer bzw. sehr kleiner extremer Übereinstimmungsscores als normal bzw. gestört klassifiziert werden. Neben der Typentheorie betrachtet McQuitty (1973) auch noch konkurrierend eine kognitive Frustrationstheorie, um Gestörte und Normale zu trennen.

Die in McQuitty (1963 a) eingeführte Rangordnungstypenanalyse hat den Nachteil, daß sie sehr strenge Anforderungen stellt, wenn Individuen zu einem Typ klassifiziert werden sollen. Falls die Daten fehlerbehaftet sind, werden

nur wenige Individuen die Bedingungen erfüllen und die Klassifikation wird bald abbrechen. In McQuitty (1971 a) wird deshalb eine *abgeschwächte Rangordnungstypenanalyse* vorgeschlagen, bei der ein Typ eine solche Klasse von Individuen darstellt, daß ein Individuum weder in der zugehörigen Zeile noch Spalte der spaltenrangtransformierten Interassoziationsmatrix einen Rang aufweist, der unterhalb eines vorgegebenen Minimums liegt. Das Minimum ist der minimale Wert, bei dem sich noch eine in sich konsistente Kategorie ergibt. Dabei wird das Kriterium für eine Typenzugehörigkeit sukzessiv herabgesetzt, bis alle Individuen klassifiziert sind oder es sich zeigt, daß sie nicht zu klassifizieren sind.

Während einige Methoden der hierarchischen Klassifikation nur die höchsten Assoziationsindizes verwenden, benutzen andere Methoden alle Indizes. Die *reliable und valide hierarchische Klassifikation* von McQuitty & Frary (1971) verwendet die Teilmenge von Assoziationsindizes, die die reliabelste und valideste Lösung liefert. Ausgangspunkt sind mehrere Definitionen eines *reinen Typs*. Ein reiner Typ kann dadurch definiert sein, daß jedes Individuum der Kategorie jedem anderen Individuum der Kategorie ähnlicher ist als irgendeinem Individuum einer anderen Kategorie. Man spricht hier von einem *quadratischen Typ*, da er einer quadratischen Teilmatrix in der spaltenrangtransformierten Interassoziationsmatrix entspricht. Eine andere Definition des reinen Typs besagt, daß jedes Individuum der Kategorie alle Charakteristika eines Musters enthält und kein Individuum außerhalb der Kategorie alle Charakteristika dieses Musters enthält. Dieses wird als *Identitätstyp* bezeichnet. Schließlich kann ein reiner Typ dadurch definiert sein, daß jedes Individuum innerhalb der Kategorie einem anderen Individuum der Kategorie am ähnlichsten ist. Dieses ist ein sogenannter *Verlängerungstyp*. Falls nur einige unpassende Assoziationsindizes verhindern, daß einer der obigen Typen resultiert, so bezeichnet man diese als *Flecken* und spricht von einem *gefleckten Typ*. Genügend große Flecken können verwendet werden, um Nichtmitglieder aufzudecken. Reine Typen sind nur in der Theorie anzutreffen. Ihre empirischen Korrelate heißen *reale Typen* und werden als gefleckte Typen identifiziert. Die Fleckenanzahl und die Nähe zum reinen Typ können jeweils bestimmt werden. Das Verfahren beruht auf einer sukzessiven Berücksichtigung der größten Assoziationsindizes.

Anhand eines speziellen Datensatzes mit komplizierter Zusammenhangsstruktur vergleicht McQuitty (1971 b) die iterative Interspaltenkorrelationsanalyse nach McQuitty & Clark (1968a), die hierarchische Klassifikation durch multiple Verbindungen (McQuitty, 1970a), die abgeschwächte Rangordnungstypenanalyse von McQuitty (1971 a) sowie die hierarchische Klassifikation durch Mehrstufenreziprozität von McQuitty (1970b). Die Absicht ist dabei, eine neue Methode zu entwickeln, die die Vorteile der älteren Methoden besitzt und gleichzeitig ihre Nachteile vermeidet.

Da die meisten hierarchischen Klassifikationsmethoden wegen des Rechenaufwandes auf relativ kleine Interassoziationsmatrizen beschränkt bleiben müssen, geben McQuitty & Koch (1975a) ein Verfahren an, das auch noch für tausend Individuen anwendbar ist. Es werden dabei *dyadische Typen* betrachtet, die nur aus einem reziproken Paar bestehen und *Typen höherer Ordnung*. Diese ergeben sich aus dyadiischen Typen, indem man bei diesen jeweils einen Paarling unberücksichtigt läßt und damit die Möglichkeit neuer mit dem Ausgangstyp assoziiierter dyadiischer Typen eröffnet. Es werden drei Versionen der Methode vorgestellt. Die Verfahren beginnen damit, daß in der Interassoziationsmatrix in jeder Spalte der höchste Index unterstrichen wird. Wenn zwei Individuen reziprok sind, so wird beim *Konzentrationsclustern* dasjenige mit der geringeren Anzahl der Zeilenunterstreichungen eliminiert. Beim *Zerstreuungsclustern* wird hingegen in diesem Fall das Individuum mit der höheren Anzahl von Zeilenunterstreichungen herausgenommen. Beim *Medianclustern* schließlich wird das Individuum mit der größeren Anzahl von Unterstreichungen bei der einen Hälfte der Paare und das Individuum mit der kleineren Anzahl bei der anderen Hälfte der Paare eliminiert.

Ein noch handlicheres Verfahren wird in McQuitty & Koch (1975 b) entwickelt. Ausgangspunkt ist die Definition eines Typs als einer Kategorie von Individuen, in der jedes Individuum der Kategorie einem oder mehreren Individuen der Kategorie am ähnlichsten ist. Das Verfahren beginnt damit, den höchsten Index in der Interassoziationsmatrix aufzusuchen. Die zugehörigen zwei Individuen werden zusammengefaßt, eines per Zufall ausgewählt und die zugehörige Zeile und Spalte gestrichen. Mit der reduzierten Matrix wird genau so verfahren, bis alle Individuen klassifiziert sind. Jedesmal, wenn zwei Individuen zusammengefaßt werden, bringen sie in die neue Kategorie alle ihnen bisher schon zugeordneten Individuen mit ein.

In Zusammenhang mit den Methoden des hierarchischen Konzentrationsclusterns (McQuitty & Koch, 1975 a), des *hierarchischen Größtes-Element-Clusterns* (McQuitty & Koch, 1975 b) und der elementaren Verbindungsanalyse (McQuitty, 1957 b) wird in McQuitty & Koch (1976) die Methode des *hierarchischen Größtes-Spaltenelement-Clusterns* entwickelt. Hier wird ein Typ als eine Kategorie angesehen, bei der für jedes Individuum der Kategorie ein anderes Individuum in der Kategorie vorhanden ist, das ihm am ähnlichsten ist. Das Verfahren isoliert in jeder Spalte ein größtes Element als wesentlich und eliminiert die anderen Daten als irrelevant. Je nach der Behandlung von Bindungen werden zwei Versionen unterschieden. Bei der einen Version werden weniger Bindungen als bei anderen Verfahren erzeugt. Falls Bindungen auftreten, werden diese besonders berücksichtigt. Bei der zweiten Version werden alle Bindungen vor der Analyse entfernt.

Von dem Paradoxon Meehls (1950) ausgehend unterscheidet McQuitty (1976) zwischen *additiver* und *konfiguraler Varianz* von Testitems. Die Faktorana-

lyse berücksichtigt nur die additive Varianz und läßt den Varianzanteil, der durch Interaktionen höherer Ordnung verursacht wird, unberücksichtigt. Die Summe beider Varianzanteile wird als *kombinierte Varianz* bezeichnet und mit der sogenannten **umfassenden Analyse der Testitemvarianz** untersucht. Ausgangspunkt ist die Unterscheidung von *Persönlichkeitstypen* und *Testitemtypen*. Ein Persönlichkeitstyp ist eine solche Kategorie, bei der ein Individuum aus der Kategorie alle oder fast alle Persönlichkeitscharakteristika eines Musters von Charakteristika besitzt, wobei dieses für kein Individuum außerhalb der Kategorie gilt. Ein Testitemtyp ist analog eine solche Kategorie, bei der ein Individuum aus der Kategorie bezüglich fast aller Antwortalternativen ein bestimmtes Muster produziert, während dieses für kein Individuum außerhalb der Kategorie gilt. Es werden drei Versionen der Methode unterschieden, wobei Version I mehr Zufallsvarianz in den Daten gestattet als Version II. Version III setzt nicht soviele perfekte typologische Antworten wie die beiden anderen Versionen voraus, verlangt dafür aber weniger Zufallsantworten für die Individuen, die gegen Ende der Analyse isoliert werden. Die Methode gestattet es, zwischen Testitemtypen zu unterscheiden, die Persönlichkeitskonstrukte widerspiegeln und solchen, die dieses nicht tun. Ferner werden die Testitems aufgedeckt, mit deren Hilfe lineare und konfigurale Typen isoliert werden können, wobei *lineare Typen* dadurch gekennzeichnet sind, daß man sie einem linearen Kontinuum zuordnen kann.

Wenngleich viele der musteranalytischen Methoden von McQuitty auch bei größeren Stichproben im Prinzip ohne Rechner durchgeführt werden können, so trifft dieses in der Praxis meist nur dann zu, falls die Interassoziationsmatrix gegeben ist. Von Lee (1977) wird angegeben, wie man aus Q-Sort-Daten (Stephenson, 1953) auf elementare Weise den Produktmomentkorrelationskoeffizienten als Grundlage einer Interassoziationsmatrix erhält.

### 3. Die Konfigurationsfrequenzanalyse von Lienert

Die *Konfigurationsfrequenzanalyse* (KFA) stellt den Versuch dar, mit Hilfe von statistischen Definitionen und Verfahren *Typen* und *Syndrome* durch Abhängigkeitsstrukturen in mehrdimensionalen Kontingenztafeln zu definieren und nachzuweisen. Ausgangspunkt ist ein Vortrag von Lienert aus dem Jahre 1968 (Lienert, 1969a). Es werden  $t$  binäre Variablen (Items) betrachtet, die zu  $2^t$  verschiedenen Mustern führen. Für eine Stichprobe von  $N$  Individuen wird ausgezählt, wie oft jedes Muster aufgetreten ist. Für jedes Muster bestimmt man den bedingten Erwartungswert  $e$  für die Auftretenshäufigkeit unter der Nullhypothese totaler Unabhängigkeit der Items bei festgehaltenen eindimensionalen Randsummen. Dazu bildet man das Produkt der betreffenden Randsummen und dividiert es durch  $N^{t-1}$ . Falls  $o$  die beobachtete Häufigkeit für ein Muster ist, vergleicht man die Chiquadratkompone (o-e)<sup>2</sup>/e

mit einem geeignet gewählten Quantil der Chiquadratverteilung mit einem Freiheitsgrad und beurteilt danach, ob man die Häufigkeit des Musters als überzufällig ansehen will. Muster, auch als *Konfigurationen* bezeichnet, die sich in höchstens einem Item unterscheiden, können *agglutiniert*, d.h. zu komplexeren Klassen zusammengefaßt werden. Nur solche Muster, die sich als überzufällig erwiesen haben und sich in wenigstens einem Item unterscheiden, werden als distinkte Klassen angesehen. Die *Effizienz* einer Klassifikation wird gemessen durch das Verhältnis der Anzahl der klassifizierten Individuen zu der Anzahl aller Individuen.

In Lienert (1969 a) wird darauf hingewiesen, daß frühere Klassifikationsansätze allein auf Double-Korrelationen, d.h. Zusammenhängen zwischen je zwei Variablen beruhen, während die KFA auch Interaktionen höherer Ordnung berücksichtigt. In Lienert (1970) werden die Ergebnisse eines LSD-Versuches beschrieben, bei dem keine Double-Korrelationen auftreten, dafür aber eine Tripel-Korrelation nachweisbar ist. Auf die Gefahr von fälschlicherweise angenommenen Syndromen und die Nichtaufdeckung vorhandener Syndrome, falls man nur Double-Korrelationen betrachtet, weisen auch Lange & Vogel (1965) hin. Sie geben verschiedene Situationen an, wo durch heterogene Untergruppen Syndrome vorgetäuscht bzw. verschleiert werden. In der KFA hofft man, solche Heterogenitäten in den Items berücksichtigt zu haben.

Einen Vergleich der KFA mit anderen typenanalytischen Verfahren geben Lienert & Kerekjarto (1969). Sie diskutieren die Möglichkeiten, depressive Symptome zu Syndromen und depressive Patienten zu Typen zu klassifizieren. Als Möglichkeiten der apriorischen Symptomklassifikation werden die psychologische Klassifikation, die psychopathologische Klassifikation, die semantische Klassifikation, die topologische Klassifikation, die systematische Klassifikation, die meßtheoretische Klassifikation und die methodologische Klassifikation diskutiert. Für die aposteriorische Klassifikation wird unterschieden zwischen Clusteranalyse, Faktorenanalyse, Musteranalyse und Konfigurationsfrequenzanalyse.

In Lienert (1969b) werden Anwendungsmöglichkeiten der KFA in der Diagnostik vorgeschlagen. Dabei werden überzufällig häufige Konfigurationen als Anzeichen von Diagnoseklassen gewertet. Um die Validität von Syndromen zu überprüfen, wird die relative Häufigkeit der Probanden mit der kritischen Konfiguration  $k$  in einer Kriteriumsgruppe ( $p_{kC}$ ) und in einer Vergleichsgruppe ( $p_{kV}$ ) bestimmt und der *Validitätskoeffizient*

$$r_{kC} = (p_{kC} - p_{kV}) / (p_{kC} + p_{kV})$$

zur Beurteilung der Validität berechnet. Die Bedeutsamkeit von  $r_{kC}$  wird durch Vergleich mit einem Chiquadratquantil beurteilt. Eine Version des Verfahrens für symptomreduzierte Konfigurationen wird für unzureichend beschriebene Patienten und für zu kleine Stichprobenumfänge empfohlen.

Die konfigurale Validierung soll nach Lienert (1973) einer *konfiguralen Skalierung* stets vorausgehen. Für die letztere wird der Anteil der Probanden der Kriteriumsgruppe geschätzt durch

$$p_k = (f_{kC} \pm 0.5) / (f_{kC} + f_{kV}),$$

wobei  $f_{kC}$  und  $f_{kV}$  die entsprechenden absoluten Häufigkeiten bedeuten und die zentralisierende Stetigkeitskorrektur von 0.5 im Zähler die Werte  $p_k = 0$  und  $p_k = 1$  verhindert. Für  $f_{kC} = f_{kV}$  ist die Korrektur zu unterlassen. Die Werte  $p_k$  werden über eine Standardnormalverteilung in z-Skalenwerte transformiert.

Die eigentliche Entwicklung der KFA beginnt mit dem Artikel Lienert (1971 a, nachgedruckt in Krauth & Lienert, 1973 b, Kapitel 1). Nach einer Erläuterung des Verfahrens der KFA an dem LSD-Beispiel von Lienert (1970) wird ein Typ für qualitative Merkmale definiert als ein Muster von Eigenschaften, das von anderen Mustern unterschieden werden kann, wobei es signifikant häufiger beobachtet wird als dem Erwartungswert unter der Nullhypothese der totalen Unabhängigkeit entsprechen würde. Dieses ist im ersten Teil die lexikalische Typendefinition von English & English (1958, S. 567), während der zweite Teil ein statistisches Moment in die Definition einbringt. Bei der Übertragung von qualitativen auf quantitative Merkmale wird die Definition Rohrachers (1975, S. 14), nach der ein Typ eine durch einen bestimmten Merkmalskomplex charakterisierte Gruppe von Individuen ist, ergänzt durch die Forderung, daß die Einzelmerkmale überzufällig oft überdurchschnittlich stark ausgeprägt sein müssen.

Neben den Typen werden auch sogenannte *Antitypen* betrachtet, die überzufällig selten auftreten unter der Nullhypothese der totalen Unabhängigkeit der Variablen. Es wird darauf hingewiesen, daß Merkmale assoziiert sein können, ohne daß Typen vorliegen und umgekehrt. Um die Stärke der Ausgeprägtheit eines Typs zu messen, wird ein sogenannter *Prägnanzkoeffizient* definiert (unterschiedlich in Lienert (1971 a) und Krauth & Lienert (1973 b)). Als Vorteile der KFA werden angesehen die Typendefinition, die nicht nur Ähnlichkeit, sondern Identität der Merkmalsmuster voraussetzt, und die Begründung durch das Modell der mehrdimensionalen Kontingenzanalyse. Als Nachteile werden angeführt die große Zahl der benötigten Individuen und die Notwendigkeit typologisch relevanter und genügend reliabler Merkmale.

Die KFA in der beschriebenen Form ist als ein *heuristisches Hilfsmittel* anzusehen, um Typen, die sich als Inhomogenitäten in einem gegebenen Datensatz abzeichnen, aufzufinden. Sie ist aber nicht in der Lage, Typen nachzuweisen.

Eine inferenzstatistische Begründung der KFA zum Nachweis von Typen wird in Krauth & Lienert (1973 a) gegeben. Neben den schon erwähnten asymptotischen Chiquadrattests zum Nachweis von Typen bzw. Antitypen treten auch für kleine N anwendbare Binomialtests und deren Normalapproximationen auf.

nen. Die Wahrscheinlichkeiten für die Binomialtests werden aus den eindimensionalen Randwahrscheinlichkeiten durch  $e/N$  geschätzt. Da dadurch die Anpassung nur verbessert werden kann, handelt es sich um ein konservatives Verfahren. Bei  $t$  Symptomen sind  $2^t$  simultane abhängige Binomialtests durchzuführen. Um die Fehlerwahrscheinlichkeit 1. Art unter Kontrolle zu halten, werden die einzelnen Tests mit dem adjustierten

$$\alpha^* = \alpha/2^t$$

beurteilt, wobei  $\alpha$  die vorher festgelegte Irrtumswahrscheinlichkeit ist. In Krauth & Lienert (1973 b, Kapitel 2) werden zusätzlich Binomialtests für die Fragestellung entwickelt, ob die  $k$  größten vorgefundenen Konfigurationshäufigkeiten noch mit der Nullhypothese der totalen Unabhängigkeit verträglich sind. Eine Darstellung der KFA sowie die Beschreibung eines FORTRAN-Programmes für diese Methode geben Roeder (1974) sowie Steinhäusen & Langer (1977, S. 148-156).

Neben den in den methodischen Arbeiten zur KFA enthaltenen Anwendungsbeispielen ist das Verfahren mittlerweile in verschiedenen Bereichen angewandt worden. So wird die KFA verwendet von Lienert & Matussek (1971), um Depressionssyndrome aufzudecken, von Müller, Ruppen, Baumann & Angst (1972), um Drogentypen zu finden, von Gloning, Quatember & Lienert (1972), um Aphasikertypen zu identifizieren, von Masendorf & Roeder (1974) sowie von Roeder & Masendorf (1979), um Leistungsausfalltypen bei Schülern zu isolieren, von Feger (1978, Kapitel 7), um Typen des Konflikterlebens zu erkennen und schließlich von Silbereisen, Oesterreich & Leitner (1977), um Typen von Sozialhilfeempfängern zu identifizieren.

In Lienert (1971 b) (nachgedruckt in Krauth & Lienert, 1973 b, Kapitel 3) wird die KFA zu der *hierarchischen Konfigurationsfrequenzanalyse* erweitert. Beabsichtigt ist hiermit die Elimination all derjenigen Variablen, die für die Typenbildung irrelevant sind. Um dieses Ziel zu erreichen, wird eine einfache KFA sowohl für die ursprüngliche Menge von Variablen als auch für jede Teilmenge der Variablen durchgeführt. Es wird empfohlen, diejenige KFA für die Interpretation näher in Betracht zu ziehen, für die das Gesamtchiquadrat, d.h. die Teststatistik für das Testen der totalen Unabhängigkeit der betroffenen Variablen, die kleinste Überschreitungswahrscheinlichkeit aufweist. Für eine Teilmenge mit  $k$  Variablen ist dabei mit  $2^k-k-1$  Freiheitsgraden der Chiquadratverteilung zu rechnen. Das Gesamtchiquadrat ergibt sich als Summe der Chiquadratkponenten  $(o-e)^2/e$  für alle betrachteten Konfigurationen. Von Rauchfleisch (1974) wird die hierarchische KFA dazu verwendet, Typen von verwahrlosten Jugendlichen aufzudecken, unter Berücksichtigung von Frustrations- und Intelligenzvariablen, während Osselmann (1979) mit ihr Depressionstypen identifiziert.

Auf der Grundlage einer Beschreibung der KFA und der hierarchischen KFA durch Lienert & Krauth (1975a) erstellten Morris (1976) und Schlattmann & Wildgrube (1979) FORTRAN-Programme.

Für den speziellen Fall, daß die Variablen die Änderung von anderen Variablen messen, lassen sich KFA und hierarchische KFA zur Identifikation von *Veränderungstypen* verwenden. So suchen Masendorf, Roeder & Kretschmann (1976) nach Änderungstypen bei Schulleistungsänderungen. Eine andere Form der Veränderungsmessung wird in Lienert (1971 d) vorgeschlagen, wo für die Werte vor der Behandlung und nach der Behandlung getrennt je eine KFA durchgeführt und die Veränderung der Typenstruktur registriert wird. Andere Verfahren zur Veränderungsmessung mit Hilfe der KFA werden noch besprochen.

Von der hierarchischen KFA werden jeweils nur Typen mit einer konstanten Anzahl von Merkmalen isoliert. Gelegentlich ist es aber zweckmäßig, eine variierende Anzahl typenerzeugender Merkmale anzunehmen. So wird man zwei überfrequentierte Konfigurationen, die sich nur in einem Merkmal unterscheiden, zu einer Konfiguration, die das kritische Merkmal nicht mehr enthält, zusammenfassen. Eine solche *agglutinierende Konfigurationsfrequenzanalyse* erlaubt die Bildung von *Agglutinationstypen* bei fehlenden oder schwach ausgeprägten Konfigurationstypen, die Bildung prägnanter Agglutinationstypen aus weniger prägnanten Solitärtypen, die Reduzierung der Typenzahl und eine einfachere Interpretation von Typen und führt zu einer Erhöhung der Erwartungswertschätzungen.

Sowohl die hierarchische KFA als auch die agglutinierende KFA sind heuristische Suchverfahren. Nachdem man mit ihrer Hilfe Hypothesen über vermutete Typen generiert hat, kann man diese an einer Kreuzstichprobe inferenzstatistisch überprüfen. Diese Überprüfung mit simultanen Binomialtests wird in Krauth & Lienert (1973 b, Kapitel 4) diskutiert.

Eine Modifikation der KFA für diagnostische Zwecke stellt die *Zwei- und Mehrstichproben-Konfigurationsfrequenzanalyse* von Lienert (1971 c) (nachgedruckt in Krauth & Lienert, 1973 b, Kapitel 5) dar. Dieses Verfahren kann in Kreuzvalidierungen verwendet werden, um zu überprüfen, ob die Konfigurationsverteilungen und damit etwaige Typen für zwei oder mehr Stichproben identisch sind. Eine andere Anwendung ist das Auffinden von Typen für Stichproben aus unterschiedlichen Populationen zum Zwecke der Klassifikation von Individuen mit bekannter Merkmalskonfiguration.

Bei t Merkmalen und k Populationen stellt man eine Kontingenztafel mit  $2^t$  Zeilen, die den Konfigurationen entsprechen, und k Spalten, die den Populationen entsprechen, auf. Zu jeder Population wird eine Zufallsstichprobe erhoben. Für jede Zelle der Kontingenztafel bestimmt man den bedingten Er-

wartungswert e der Zellhäufigkeit als Produkt der beiden Randsummen dividiert durch die Summe aller Stichprobenumfänge. Dann berechnet man die Chiquadratkomponenten für jede Zelle nach der Beziehung  $(o-e)^2/e$ , wobei o die Zellfrequenz ist, und bildet die Zeilensummen über diese Komponenten. Falls eine solche Zeilensumme oberhalb eines geeignet gewählten Quantils einer Chiquadratverteilung mit  $(k-1)$  Freiheitsgraden liegt, sieht man die zugehörige Konfiguration als geeignet an, zwischen den Populationen zu diskriminieren. Eine Dateninspektion ergibt dann, für welche Populationen die betreffende Konfiguration typisch und für welche sie atypisch ist. Dadurch ist es möglich, Individuen aufgrund ihrer Konfigurationen zu klassifizieren. In analoger Weise kann man hierarchische *Mehrstichproben-Konfigurationsfrequenzanalysen* und *agglutinierende Mehrstichproben-Konfigurationsfrequenzanalysen* durchführen.

Die heuristische Zwei- und Mehrstichproben-KFA wird in Krauth & Lienert (1973 b, Kapitel 6) inferenzstatistisch begründet. Die Auswertung erfolgt über Kontingenztafeln mit 2 Zeilen und k Spalten. In der ersten Zeile stehen die Frequenzen für die betrachtete Konfiguration aufgeteilt nach den k Stichproben. In der zweiten Zeile stehen die Summen der Frequenzen über alle übrigen Konfigurationen. Bei  $2^t$  Konfigurationen sind ohne vorherige Kenntnis geeigneter Hypothesen  $2^t$  simultane Chiquadrattests über  $(2 \times k)$ -Kontingenztafeln mit Alphaadjustierung durchzuführen. Speziell für die Zweistichproben-KFA mit  $k = 2$  Stichproben ergeben sich Vierfeldertafeln, die man bei kleinen Stichprobenumfängen auch mit dem exakten Test von Fisher auswerten kann. Bezuglich der exakten Tests bei mehr als zwei Stichproben vergleiche man Freeman & Halton (1951), March (1972), Krauth (1973) Boulton (1974), Stegie & Wall (1974), Stucky & Vollmar (1975), Hancock (1975), Howell & Gordon (1976), Baker (1977) und Bedeian & Armenakis (1977).

Da man mit einem  $(2 \times k)$ -Feldertest nur nachweisen kann, welche Konfigurationen diskriminieren, aber nicht, ein welcher Weise sie diskriminieren, werden wieder simultane Vierfeldertests vorgeschlagen. Eine solche Vierfeldertafel enthält die Frequenz zu einer vorgegebenen Konfiguration und einer vorgegebenen Stichprobe, die Summe der Frequenzen über alle restlichen Stichproben bei gleicher Konfiguration, die Summe der Frequenzen über alle restlichen Konfigurationen bei gleicher Stichprobe und die Ergänzung bis zur Summe aller Stichprobenumfänge. Mit einem solchen Vierfeldertest lässt sich überprüfen, ob das Vorhandensein einer speziellen Konfiguration typisch für eine bestimmte Population ist. Wegen der i.a. großen Anzahl simultaner Vierfeldertests wird es zweckmäßig sein, an einer Pilotstudie relativ wenige Hypothesen zu formulieren, die an einer Hauptstudie überprüft werden können.

In Baumann (1973) werden die KFA, die hierarchische KFA und die Zwei-stichproben-KFA mit anderen Typenanalyse-Verfahren verglichen. Es wird dabei sowohl ein theoretischer Vergleich, die Ziele und Voraussetzungen der

Verfahren betreffend, durchgeführt als auch eine Plasmodenstudie, d.h. ein numerischer Vergleich der verschiedenen Verfahren an einem bekannten Datensatz. Anstelle der Vierfeldertests schlägt Roeder (1976) simultane Konfidenzintervalle vor, um in der Mehrstichproben-KFA Frequenzen von Konfigurationen zu vergleichen. Von Schott & Schott (1976) wird eine Zweistichproben-KFA dazu verwendet, Studienanfänger der Medizin, die entweder aufgrund der guten Abiturnote oder aufgrund einer langen Wartezeit einen Studienplatz erhielten, nach ihrer psychosozialen Struktur zu differenzieren.

Ein alternativer Ansatz zur Mehrstichproben-KFA als Klassifikationsverfahren wird von Hommers (1978) vorgeschlagen. Er nimmt an, daß für jeden Probanden Wiederholungen eines Antwortvektors vorliegen. Falls diese Wiederholungen, z.B. durch verschiedene Probleme gekennzeichnet, als *unabhängig* voneinander angesehen werden dürfen, ergibt sich für die Verteilung des Antwortvektors eine Multinomialverteilung. Die Parameter dieser Verteilung werden aus der Gesamtstichprobe geschätzt. Mit Hilfe des Satzes von Bayes (vgl. Kap. 4 des Bandes „Hypothesenprüfung“ aus der Enzyklopädie der Psychologie) werden dann Individuen aufgrund ihres Vektors von Konfigurationshäufigkeiten einer bestimmten Population zugeordnet. Eine Anwendung des Verfahrens bei der Untersuchung der Therapiemotivation von Klienten diskutieren Steller & Hommers (1977).

Die mit der KFA nachzuweisenden Konfigurationstypen werden zurückgeführt auf Assoziationen erster Ordnung und höherer Ordnung zwischen den beteiligten Merkmalen. Die gesamte Assoziationsstruktur versucht die Assoziationsstrukturanalyse nach Lienert (1972) (nachgedruckt in Krauth & Lienert, 1973b, Kapitel 7) aufzudecken. Dieses Verfahren beruht auf der Definition von Interaktionen höherer Ordnung nach Lancaster (1951, 1969). Im ersten Verfahrensschritt wird bei t Merkmalen das Gesamtchiquadrat für den Test auf totale Unabhängigkeit der t Merkmale berechnet. Die gleiche Rechnung wird durchgeführt für alle Kombinationen von t-1 Variablen, von t-2 Variablen usw. bis zu 2 Variablen. Das Vorliegen von Double-Assoziationen überprüft man direkt an den Gesamtchiquadraten für je 2 Variablen durch Vergleich mit einem Chiquadratquantil für 1 Freiheitsgrad. Von jedem Gesamtchiquadrat für je 3 Variablen zieht man nun die zugehörigen  $\binom{3}{2} = 3$  Double-Chiquadrante ab und erhält jeweils ein Residuum, das man bei einem Freiheitsgrad auf eine Tripel-Assoziation hin überprüft. Von jedem Gesamtchiquadrat für je 4 Variablen zieht man die zugehörigen  $\binom{4}{2} = 4$  Tripel-Chiquadrante und  $\binom{4}{3} = 6$  Double-Chiquadrante ab und überprüft das Residuum bei einem Freiheitsgrad auf eine Quadrupel-Assoziation. Bei t Variablen erhält man bei Fortsetzung des Verfahrens  $\binom{t}{2}$  Double-Chiquadrante,  $\binom{t}{3}$  Tripel-Chiquadrante, . . . ,  $\binom{t}{t} = 1$  t-faches Chiquadrat, insgesamt  $2^t - t - 1$  verschiedene Assoziations-Chiquadrante. Die verschiedenen sich zeigenden Assoziationen kann man durch *Assoziationsgraphen* veranschaulichen, in denen die Merkmale durch Knoten und die Assoziationen durch Kanten symbolisiert sind.

Die Erweiterung von KFA und Assoziationsstrukturanalyse von Merkmalen mit nur zwei Ausprägungen auf Merkmale mit mehr als zwei und u.U. unterschiedlich vielen Ausprägungen bereitet keine prinzipiellen Schwierigkeiten. Für die Assoziationsstrukturanalyse wird diese Erweiterung als *Kontingenzstrukturanalyse* bezeichnet und in Lienert & Krauth (1973a) (nachgedruckt in Krauth & Lienert, 1973b, Kapitel 9) dargestellt.

So anschaulich die Assoziationsstrukturanalyse auch ist, bei der Assoziationen höherer Ordnung heuristisch erklärt werden durch den ‚Rest‘, der nach Abzug der Interaktionen niedrigerer Ordnung noch verbleibt, so schwerwiegende Bedenken müssen aus theoretischer Sicht erhoben werden. Wie in Krauth & Lienert (1973b, Kapitel 8) dargelegt, ist die Interpretation der Interaktionen höherer Ordnung nach Lancaster (1951, 1969) nicht leicht und die Zerlegbarkeit der Chiquadrate, wie sie oben ausgenutzt wurde, ist nur asymptotisch und nur unter der Nullhypothese der totalen Unabhängigkeit der Merkmale gegeben. Will man dennoch eine solche Analyse durchführen, so sind entsprechende Assoziationen inferenzstatistisch durch simultane Chiquadrattests mit adju- stiertem Alpharisiko abzusichern.

Eine Kurzdarstellung von KFA und Assoziationsstrukturanalyse geben Lienert & Krauth (1976), während Gebert (1978) mit Hilfe der Assoziationsstrukturanalyse zeigt, daß in den Beispielen, für die in der Literatur latente Klassenanalysen gerechnet wurden, nur Interaktionen erster Ordnung auftreten.

Von Baumann, Gebert & Lienert (1973) wird die Assoziationsstrukturanalyse zum Nachweis von Depressionssyndromen verwendet und ein Vergleich der Ergebnisse mit denen einer parallel durchgeführten Faktorenanalyse vorgenommen. In Rey, Klug & Welz (1978a, 1978b) wird eine Stichprobe von psychiatrischen Patienten mit einer Normalstichprobe mittels einer Zweistichproben-KFA verglichen. Anschließend wird die Stichprobe der psychiatrischen Patienten mit Hilfe einer KFA, einer hierarchischen KFA und einer Assoziationsstrukturanalyse auf das Vorhandensein von Typen hin analysiert.

Aufgrund der theoretischen Schwierigkeiten, die bei der Interpretation der Ergebnisse von Assoziationsstrukturanalysen auftreten, schlagen Krauth & Lienert (1974) als alternatives Verfahren die sogenannte *Interaktionsstrukturanalyse* vor. Dieser liegt ein anderes Konzept von Interaktionen höherer Ordnung zugrunde als der Assoziationsstrukturanalyse und anderen bisher bekannten Verfahren. Alle diese Verfahren gehen z.B. bei drei Merkmalen A, B und C von einer einzigen Interaktion 2. Ordnung aus. Das neue Verfahren definiert drei Interaktionen 2. Ordnung, jeweils zwischen einem Merkmal und dem restlichen Merkmalspaar, formal bezeichnet mit A.BC, B.AC und C.AB. Z.B. testet man, ob eine Interaktion A.BC vorliegt, indem man eine Kontingenztafel aufstellt, in der die Zeilen den Ausprägungen von A und die Spalten den Ausprägungskombinationen von BC entsprechen, und den Chiquadrat-test

auf Unabhängigkeit durchführt. Entsprechend erhält man bei vier Merkmalen A, B, C und D 6 Interaktionen 1. Ordnung, z.B. A.B, 12 Interaktionen 2. Ordnung, z.B. A.BC, und 7 Interaktionen 3. Ordnung, z.B. A.BCD oder AB.CD. Jede dieser Interaktionen kann mit Hilfe einer zweidimensionalen Kontingenztafel leicht interpretiert und getestet werden. Die inferenzstatistischen Aspekte dieses Verfahrens werden in Krauth & Lienert (1973b) ange- sprochen.

Von Wermuth (1976) wird anhand einiger fiktiver Datensätze illustriert, daß KFA, hierarchische KFA und Interaktionsstrukturanalyse nicht in allen Fällen die gesamte Abhängigkeitsstruktur aufdecken, falls etwa bedingte Abhängigkeiten vorliegen. Derartige Abhängigkeitsstrukturen können u.a. mit Hilfe von multiplikativen Modellen (Wermuth, 1978), die wiederum Sonderfalle der loglinearen Modelle sind, aufgeschlüsselt werden. Eine derartige mehr oder weniger vollständige Analyse der Abhängigkeitsstruktur, wie sie z.B. ähnlich auch in Viktor (1972) oder Enke (1975) angestrebt wird, dürfte allerdings in den meisten Fällen, wo es um den Nachweis und die Erklärung von Typen und Syndromen geht, zu aufwendig sein und in der Praxis zu Interpretations- schwierigkeiten führen.

Eine Weiterentwicklung der Interaktionsstrukturanalyse ist die *Prädiktions- Konfigurationsfrequenzanalyse* von Lienert & Krauth (1974a). Hier geht es darum, einen Prädikanden Y aus ( $t - 1$ ) Prädiktoren  $X_1, \dots, X_{t-1}$  vorherzusagen. Dazu beurteilt man die Interaktion  $Y \cdot X_1, \dots, X_{t-1}$ . Ist man an einer differentiellen Prädiktion interessiert, so führt man mit der sogenannten *differentiellen Prädiktions-Konfigurationsfrequenzanalyse* eine KFA über die Zellen der zu der obigen Interaktion gehörigen Kontingenztafel durch, um *Prädiktionstypen* zu finden. Wenn man etwa Schulerfolg vorhersagen will, so kann man aufbauend auf biosozialen und Testprädiktoren nach Erfolgs- und Mißerfolgstypen suchen. Von Lienert & Wolfrum (1979) wird die Prädiktions-KFA verwandt, um die Wirkungen verschiedener Psychotherapieformen in Abhängigkeit von der sozioökonomischen und anamnestischen Zugehörigkeit der Patienten zu beurteilen. Krüger (1979) vergleicht die Prädiktions-KFA mit der sogenannten *hierarchischen Rangvarianzanalyse* von Schulze (1978) und schlägt für die Auswahl von optimalen Prädiktorensets eine Assoziations- strukturanalyse vor. Eine Erweiterung des Verfahrens auf zwei Prädikanden und eine Version für geschichtete Stichproben geben Heilmann, Lienert & Maly (1979).

Auf die Möglichkeit, mittels KFA Veränderungstypen zu suchen, weist schon Lienert (1971b) hin. In Lienert & Krauth (1973b) (nachgedruckt in Krauth & Lienert, 1973b, Kapitel 10) werden *Symptomverschiebungstypen* gesucht, indem man Variablen definiert, die die Veränderung der Symptome beschreiben, und diese einer KFA unterzieht. In ähnlicher Weise kann man die Änderung von Profilen erfassen und sich für *Profiländerungstypen* interessieren. Falls

nicht alle Merkmale Verschiebungen ausdrücken, werden die resultierenden Typen als *Kombinationstypen* bezeichnet.

Beobachtet man t Merkmale simultan an einem Individuum, so spricht man von einem *Profil*. Beobachtet man dasselbe Merkmal sukzessiv t Male an einem Individuum, so spricht man von einer *Konstellation*. In Lienert & Krauth (197%) (nachgedruckt in Krauth & Lienert, 1973b, Kapitel 11) wird die KFA dazu verwandt, *Konstellationstypen* aufzudecken. Erhebt man Konstellationen in zwei aufeinanderfolgenden Zeitabschnitten, z.B. in einem Überkreuzungsplan, so kann man die Veränderung der Konstellationen durch Differenzbildung in den einzelnen Komponenten der Konstellationen messen und mittels KFA *Konstellationsänderungstypen* identifizieren. Betrachtet man gleichzeitig die Konstellationen eines Zielmerkmals und die Konfigurationen von Bedingungsmerkmalen, so lassen sich mit Hilfe einer Interaktionsstrukturanalyse sowohl *Konstellationsvorhersagen* als auch *Bedingungsrückschlüsse* durchführen. Bei therapeutischen Prozeßerhebungen werden i. a. v Variablen zu t Zeitpunkten erhoben, so daß man bei geeigneter Betrachtungsweise von *Profilkonstellationen* sprechen kann. Man kann auch Versuchspläne mit Profilkonstellationsänderungen betrachten. Die KFA gestattet es dann, sowohl nach *Profilkonstellationstypen* als auch nach *Profilkonstellationsänderungstypen* zu suchen.

Die Auswertung multivariater Versuchspläne mittels der Verfahren der KFA wird in Lienert & Krauth (1973d) angesprochen. Insbesondere werden Versuchspläne ohne Meßwiederholung diskutiert. Bei den unifaktoriellen Versuchsplänen werden der *univariate*, *bivariate* oder *multivariate Mediantest* bzw. die korrespondierende Zwei- oder Mehrstichproben-KFA eingesetzt. Die bifaktoriellen und multifaktoriellen Versuchspläne werden mit der Interaktionsstrukturanalyse oder auch mit der Assoziationsstrukturanalyse ausgewertet.

Eine systematische Darstellung der Auswertung multivariater klinischer Untersuchungspläne unter Verwendung der KFA wird in Lienert & Krauth (1974b, 1974c) gegeben. Die Pläne werden klassifiziert nach der Anzahl der Faktoren und ihrer Stufen, nach der Anzahl der Observablen und ihrer Ausprägungen und danach, ob es sich um Pläne mit oder ohne Meßwiederholung handelt. Für afaktorielle Pläne ohne Meßwiederholung wird eine KFA und für entsprechende faktorielle Pläne eine Interaktionsstrukturanalyse vorgeschlagen. Bei Plänen mit einfacher Meßwiederholung werden die Differenzen mittels KFA und Interaktionsstrukturanalyse beurteilt. Bei Plänen mit mehrfacher Meßwiederholung werden Polynome an die Meßwertvektoren angepaßt und die Vektoren der Polynomkoeffizienten mittels KFA und Interaktionsstrukturanalyse ausgewertet. Eine Kurzdarstellung für die Auswertung von Meßwiederholungsplänen geben Lienert & Krauth (1975b).

Bei der Auswertung der Meßwiederholungspläne wird im Modell angenommen, daß die Meßwerte durch eine zeitabhängige Regressionsfunktion beschrieben werden, wobei eine Korrelation aufeinanderfolgender Meßwerte zugelassen ist. Die Regressionsfunktion wird dann durch ein Polynom approximiert. Statistische Tests für den Gruppenvergleich von Individuen mit solchen *Verlaufskurven* werden in Krauth (1973) entwickelt. Ist man weniger an Gruppenvergleichen als vielmehr an *Verlaufstypen* interessiert, so kann man eine KFA über die dichotomisierten Polynomkoeffizienten bzw. über die Vorzeichenfolgevektoren durchführen (Krauth & Lienert, 1975). In Krauth & Lienert (1978) werden an zwei Stichproben von Lernkurven orthogonale Polynome angepaßt und eine Zweistichproben-KFA über die dichotomisierten Polynomkoeffizienten durchgeführt.

In Bierschenk & Lienert (1977) und Bartoszyk & Lienert (1978) werden anhand einer Stichprobe von Probanden Verlaufskurventypen gesucht. Dazu wird für je zwei zeitlich aufeinanderfolgende Meßwerte das Vorzeichen der Differenz bestimmt und über die  $2^{t-1}$  Vorzeichenfolgevektoren bei t Zeitpunkten eine KFA durchgeführt. Die angenommene Gleichverteilung der Muster unter der Hypothese der Verlaufshomogenität und Trendlosigkeit beruht allerdings auf einem Irrtum, da sie z.B. der Annahme einer Gleichverteilung über alle möglichen Anordnungen der Meßwerte widerspricht. Da auch die Nullhypothese der Gleichverteilung der Rangmuster zu Interpretationsschwierigkeiten führen kann, scheint das Problem einer Typisierung von Verlaufskurven durch derartige Ansätze noch ungelöst zu sein.

#### 4. Andere typenanalytische Ansätze

##### 4.1 Die Musterähnlichkeitsanalyse von Cattell

In Cattell (1943) wird ein später als *approximative Abgrenzungsmethode* bezeichnetes Verfahren beschrieben, mit Hilfe dessen Variablen im Vorfeld einer Faktorenanalyse zu Clustern gruppiert werden können. Dieses Verfahren wird in Cattell (1944) mit drei anderen Clustermethoden verglichen, die ebenfalls von einer Korrelationsmatrix der Ausgangsvariablen ausgehen. Es handelt sich hierbei um die *Verzweigungs-Verbindungs-Methode*, die *Matrix-Diagonal-Methode* und die *Korrelations-Profil-Korrelation* von Tryon (1939). Die approximative Abgrenzungsmethode beginnt ebenso wie die Verzweigungs-Verbindungs-Methode damit, daß für jede Variable diejenigen Variablen festgestellt werden, die mit ihr die höchsten Korrelationen, sogenannte *Einfachverbindungen*, haben. Bei der Matrix-Diagonal-Methode werden die Variablen so umgeordnet, daß die Verbindungskorrelationen sich um die Hauptdiagonale der Korrelationsmatrix gruppieren, während bei Tryon's Methode die Variablen eines Clusters genügend stark positiv miteinander korrelieren müssen.

In Cattell (1949) werden verschiedene Koeffizienten definiert, um die Ähnlichkeit von Mustern zu messen, die auf intervallskalierten Variablen beruhen. Es wird ein sogenannter *Formkorrelationskoeffizient*  $r_s$  gebildet, der die Formähnlichkeit von Mustern oder Profilen messen soll. Auch die *Q-Technik* (Stephenson, 1936a, 1936b; Burt, 1937) zielt auf eine solche Formähnlichkeit von Profilen ab.  $r_s$  ist der Produktmomentkorrelationskoeffizient zwischen den beiden Profilen, wobei jedoch jede der  $n$  Profilvariablen zuvor über die Stichprobe standardisiert worden ist. Diese Standardisierung bewirkt, daß Variablen mit großer Streuung keine nur scheinbar großen Profilunterschiede vor täuschen können. Soll nicht nur die Form der Muster ähnlich sein, so wird der *Musterähnlichkeitsindex*  $r_p$  vorgeschlagen, der durch

$$r_p = (2k - \sum d_i^2) / (2k + \sum d_i^2)$$

definiert ist. Dabei ist  $k$  der Median einer Chiquadratverteilung mit  $n$  Freiheitsgraden (bei  $n$  Variablen) und  $d_i$  ist die Differenz der Standardscores für die  $i$ -te Variable. Für diskrete Variablen wird

$$r_p = (k - D/2) / (k + D/2)$$

vorgeschlagen, wobei  $D$  die Anzahl der nicht übereinstimmenden Variablen ist. Betrachtet man nur speziell eine Variable, so entfallen die Summenzeichen in  $r_p$  und man erhält den Koeffizienten der *Muster-Effekt-Ähnlichkeit*  $r'_p$ , der auf die Ursachen der Ähnlichkeit abzielt. In Cattell (1950) wird  $r_p$  benutzt, um Kulturmuster zu clustern. Einen Vergleich mit anderen Ähnlichkeitskoeffizienten unter Bezug auf DuMas (1946), Cronbach & Gleser (1953), Osgood & Suci (1952), Webster (1952), Zubin (1938) und Gaier & Lee (1953) gibt Cattell (1957). Die Verteilung von  $r_p$  unter der Annahme von normalverteilten Variablen untersucht Horn (1961).

In Cattell (1952) wird die Technik der Q-Faktorenanalyse, die auf Burt (1937) und Stephenson (1936a, 1936b) zurückgeht, als ein typenanalytisches Verfahren eingestuft und die Stellung innerhalb der anderen faktorenanalytischen Methoden geklärt. Jedoch wird in Cattell (1957, S. 408) die Q-Technik deshalb kritisiert, weil nach Verständnis des Autors Typen als Cluster und nicht als Faktoren aufzufassen sind.

Es wird (Cattell, 1957, S. 371) folgende Klassifikation zugrundegelegt: Falls die Elemente eines Musters geordnet sind, so spricht Cattell von einer *Konfiguration*, sonst von einem *Profil*. Je nachdem, ob Profile direkt bezüglich ihrer Ähnlichkeit verglichen werden oder ob über den Profilen definierte Funktionen verglichen werden, spricht Cattell von *allgemeinen* bzw. *spezifischen* Typen. Auf diesem Hintergrund vergleicht Cattell (1957, S. 374-378) den musteranalytischen Ansatz von McQuitty (1953), die latente Strukturanalyse von Lazarsfeld (1950) und Gibson (1955), die Q'-Technik von Stephenson (1953),

die kategoriale Klassifikation von Ellson (1955), die Mustervorhersage von Meehl (1950), die Musteranalyse von Lykken (1956), die manifeste Strukturanalyse von DuMas (1956) sowie die Verfahren von Horst (1954) und Lubin (1950).

In Cattell & Coulter (1966) werden Typen als Gipfel in einer multivariaten Verteilung definiert. Dabei werden zwei Typenkonzepte unterschieden. Unter einem *Homostat-Typ* wird eine Personenmenge verstanden, die durch benachbarte Punkte im Sinne eines Ähnlichkeitsmaßes im Variablenraum charakterisiert ist. Unter einem *Segregat-Typ* wird demgegenüber eine Menge von Personen verstanden, die untereinander (möglicherweise über „Zwischenträger“) verbunden d.h. ähnlich sind, aber mit Personen außerhalb des Typs nicht verbunden sind. Es werden zwei Ansätze zum Auffinden von Typen diskutiert. Bei der *interindividuellen Beziehungs-Methode* berechnet man für jedes Paar von Individuen einen Ähnlichkeitsindex, vorzugsweise  $r_p$ , und erhält damit eine sogenannte *Q-Matrix*. Neben  $r_p$  wird auch ein Koeffizient  $r_s$  (nicht zu verwechseln mit dem ebenfalls  $r_s$  genannten Formkorrelationskoeffizienten von Cattell (1949)) der *linearen Ähnlichkeit* und ein *Benachbartheitkoeffizient*  $r_n$  eingeführt. Auf die Q-Matrix wird dann die Verzweigungs-Verbindungs-Methode angewandt, um sogenannte *Phänomenal-Cluster* zu erhalten. Diese Teilkasse der Homostattypen ist allein durch die Verbindungen zwischen den Personen definiert, während die sogenannten *Nuklear-Cluster* noch weitere Struktureigenschaften besitzen müssen. Das Verfahren liefert eine Liste der Phänomenal- und der Nuklear-Cluster. Ausgangspunkt für das Auffinden von Segregattypen ist eine Matrix, die die Überlappungen zwischen den Phänomenalclustern beschreibt. Die ganze Analyse wird anwendbar durch die Entwicklung eines mit *Taxonome* bezeichneten Computerprogramms. Eine ausführliche Darstellung der Ähnlichkeitsmaße von Cattell und des Taxonome-Programms gibt Baumann (1971). Auch in Cattell, Coulter & Tsujioka (1966) sowie in Cattell (1968) werden diese Themen angesprochen.

Das Ähnlichkeitsmaß  $r_p$  mißt die Ähnlichkeit zwischen zwei Individuen. Das Problem, die Ähnlichkeit zweier Gruppen bzw. die Ähnlichkeit eines Individuums mit einer Gruppe zu bestimmen, wird in Cattell (1969) diskutiert.

## 4.2 Die Profildistanzanalyse von Sawrey, Keller und Conger

Eine Typenanalyse, die auf dem euklidischen Distanzmaß zwischen je zwei Profilen beruht, wird in Sawrey, Keller & Conger (1960) beschrieben. Der Abstand zwischen den Profilen von zwei Personen wird bestimmt, indem man für jede Variable die Differenz der Meßwerte zwischen den Personen bestimmt, diese quadriert und die Summe über die  $k$  Variablen des Profils bildet. Dieses ergibt den Abstand  $d^2$  zwischen zwei individuellen Profilen, der auch in

den Musterähnlichkeitsindex  $r_p$  von Cattell (1949) eingeht. Voraussetzung für die Berechnung von  $d^2$  ist Intervallskalenniveau der Variablen. Bei einer Stichprobe von N Personen kann man die Abstände für jedes Personenpaar in eine  $(N \times N)$ -Matrix eintragen.

In einem zweiten Schritt werden potentielle Kerngruppen gebildet. Dazu wird definiert, wie groß  $d^2$  maximal sein darf, damit noch ähnliche Profile vorliegen. Man setzt diesen Wert von  $d^2$  gleich einem Viertel der Summe über die k empirischen Varianzen, die man für jede der k Variablen über die Personen berechnen kann. Für jedes der N Profile wird dann eine Liste aller derjenigen Profile erstellt, die ihm ähnlich sind. Das Profil mit den meisten zugehörigen ähnlichen Profilen bildet mit diesen zusammen die erste potentielle Kerngruppe. Diese Profile werden gestrichen und aus dem Rest analog die zweite potentielle Kerngruppe bestimmt. Dieses Verfahren wird fortgesetzt, bis keine ähnlichen Profile mehr übrig sind.

In einem dritten Schritt werden einander unähnliche potentielle Kerngruppen bestimmt. Unähnlichkeit zwischen Profilen wird angenommen, falls  $d^2$  größer als die Summe der k empirischen Varianzen ist. Aus den Profilen, die zu potentiellen Kerngruppen gehören, wird eine Matrix von  $d^2$ -Werten aufgestellt und die Spaltensummen berechnet. Beginnend bei dem Profil mit der größten Spaltensumme werden alle Profile, die dem jeweils betrachteten Profil nicht unähnlich sind, ausgeschieden. Die so erhaltenen Kerngruppen sind einander dann alle unähnlich.

Im vierten Schritt wird für jede Kerngruppe das mittlere Gruppenprofil bestimmt. Weiter werden die möglichen Zugänge zu jeder Kerngruppe aufgelistet, d.h. solche Profile, die keinem Profil der jeweiligen Kerngruppe unähnlich sind. Nach Berechnung der  $d^2$ -Werte zwischen diesen Zugängen und den mittleren Gruppenprofilen werden die Zugänge sukzessiv aufgrund von 5 gegebenen Maximalgrenzen zu den Kerngruppen hinzugefügt, falls dieses eindeutig möglich ist. Dabei werden die Maximalgrenzen immer weiter erhöht und jedesmal für die erweiterten Gruppen neue mittlere Gruppenprofile bestimmt.

Nach Abschluß des Verfahrens erhält man Gruppen von Profilen bzw. Typen, in denen die Profile der Gruppenmitglieder einander ähnlich sind, während sie gleichzeitig den Profilen von Mitgliedern anderer Gruppen unähnlich sind.

### 4.3 Die Übereinstimmungsanalyse von Gengerelli

Im Gegensatz zur latenten Strukturanalyse (Lazarsfeld, 1950) versucht Gengerelli (1961) nicht, theoretische Strukturen, die die Daten verursachen könnten, zu finden, sondern eher eine beschreibende Klassifikation. Es werden Teil-

mengen von in gleicher Weise reagierenden Variablen gesucht und eine Möglichkeit, die gefundenen Teilmengen mit Hilfe eines Signifikanztests als tatsächlich vorhanden nachzuweisen.

Für jeden von M Patienten seien m binäre (vorhanden-nichtvorhanden bzw. 1-0) Items erhoben worden. Es soll geprüft werden, ob unter den m Items in der betrachteten Population Untermengen solcher Items existieren, die in der gleichen Richtung ausgeprägt sind. Solche zusammengehörigen Items bilden dann sogenannte *Syndrome* oder *Konstellationen*. In die Zellen einer ( $m \times m$ )-Matrix wird jeweils eingetragen, bei wie vielen von den insgesamt M Patienten beide zugehörigen Items mit 1 (Merkmal vorhanden) beantwortet werden. Um festzustellen, ob gegenseitige Übereinstimmungen bzw. Assoziationen zwischen den Items bestehen, stellt Gengerelli (1961) für jeden Tafeleintrag eine Vierfeldertafel auf. Summiert man alle Einträge in der Matrix, so erhält man die Anzahl N aller Übereinstimmungen zwischen je zwei Items, indem man die Summe der Einträge halbiert, da jeder Eintrag doppelt auftritt. Für das Matrixelement  $f_{ij}$  im Schnittpunkt der Zeile i mit der Spalte j erhält man die Spalten- bzw. Zeilensummen  $f_i$  und  $f_j$ . Die von Gengerelli (1961) aufgestellte Vierfeldertafel enthält in der ersten Zeile die Einträge  $f_{ij}$  und  $f_j - f_{ij}$  und in der zweiten Zeile die Einträge  $f_i - f_{ij}$  und  $(N - f_i) - (f_j - f_{ij})$ . Für diese Vierfeldertafel wird der übliche Chiquadratwert berechnet und mit dem 1%-Quantilwert einer Chiquadratverteilung mit einem Freiheitsgrad verglichen. (Vom statistischen Standpunkt aus wäre es möglicherweise sinnvoller, die Anzahlen  $n_i$  bzw.  $n_j$  der Patienten, die das Merkmal i bzw. das Merkmal j aufweisen, zu bestimmen und eine Vierfeldertafel zu betrachten, die in der ersten Zeile die Eintragungen  $f_{ij}$  und  $n_j - f_{ij}$  und in der zweiten Zeile die Eintragungen  $n_i - f_{ij}$  und  $(M - n_i) - (n_j - f_{ij})$  aufweist.)

Die in dieser Weise ausgezeichneten Zellen der Matrix werden markiert, und es wird überprüft, ob sich Syndrome oder Konstellationen zeigen, die in der Sprechweise der Graphentheorie durch sogenannte *vollständige Graphen* definiert sind. Z.B. bilden die Items i, j, k und l ein Syndrom, falls die Zellen (i,j), (i,k), (i,l), (j,k), (i,l) und (k,l) alle markiert sind, d.h. Chiquadratwerte oberhalb des betrachteten Quantilwerts haben. Gengerelli (1961) spricht dann auch von einem *geschlossenen* Muster, andernfalls von einem *offenen* Muster. Um solche Syndrome zu ermitteln, wird für jede Spalte der oben konstruierten Matrix die Anzahl der markierten Zellen festgestellt und zu dieser eine 1 addiert. Diese Zahl gibt die maximale Größe des Syndroms an, zu dem das Item, das die betreffende Spalte kennzeichnet, gehören kann. Man sucht die beiden Items i und j mit den höchsten Spaltensummen auf und ermittelt, ob die Zelle (i,j) markiert ist. Falls dieses der Fall ist, sucht man das Item k mit der nächsthöheren Spaltensumme und überprüft, ob (i,k) und (j,k) markiert sind. Falls dieses der Fall ist, nimmt man k zu i und j dazu. Andernfalls führt man die Vergleiche mit dem Item l mit der nächsthöheren Spaltensumme durch, usw.

Um die Zufälligkeit so gefundener Syndrome zu testen, geht Gengerelli (1961) von der Nullhypothese aus, daß alle Verteilungen von  $h$  markierten Zellen auf die  $K = m(m-1)/2$  Zellen der oberen Dreiecksmatrix ohne die Hauptdiagonale die gleiche Wahrscheinlichkeit besitzen sollen. Es wird dann die bedingte Wahrscheinlichkeit dafür berechnet, daß sich mindestens ein geschlossenes Muster aus  $d$  Items ergibt unter der Bedingung, daß  $h$  markierte Zellen unter insgesamt  $K$  Zellen vorliegen.

Die oben beschriebene Methode verwendet Gengerelli (1963) dazu, um Typen, d.h. Untergruppen in einer Population zu finden. Dabei wird ein Typ als eine Teilmenge von Punkten im Testraum, durch die die Personen gekennzeichnet werden, beschrieben derart, daß der Abstand für zwei Punkte des Typs kleiner ist als der Abstand zwischen einem Punkt des Typs und einem beliebigen Punkt außerhalb des Typs. Dadurch können Typen nur dann identifiziert werden, falls sie hinreichend weit getrennt sind. Ausgangspunkt ist hier eine  $(M \times M)$ -Matrix, in die geeignet definierte Distanzen zwischen den Personen als Punkte im Testraum eingetragen werden. Durch Festlegung einer kritischen Distanz unterscheidet man „benachbarte“ von „entfernten“ Personen und kann wieder die Übereinstimmungsanalyse durchführen.

Das von Gengerelli (1961, 1963) vorgeschlagene Verfahren kann als ein Vorfänger der auf der Theorie der Zufallsgraphen basierenden Cluster-Theorie von Ling (1972, 1973, 1975) angesehen werden.

#### 4.4 Die Ähnlichkeitsanalyse nach Lorr und McNair

Von Lorr, Klett & McNair (1963, S. 172) wird eine auf Ähnlichkeitsmaßen beruhende Typenanalyse vorgeschlagen, die sich von anderen Ansätzen (vgl. Lorr, Klett & McNair, 1963, S. 128-144; Lorr, 1966a) unterscheidet. Damit eine Typenanalyse überhaupt sinnvolle Ergebnisse liefert, fordert Lorr (1966a), daß die gefundenen Typen bei Verwendung anderer Maßskalen und anderer Stichproben replizierbar sind, daß alle Untergruppen einer gegebenen Population für eine Analyse zur Verfügung stehen und jeder identifizierte Typ durch relativ wenige Klassifikationsvariablen beschreibbar ist. Im Gegensatz zu anderen Verfahren verwendet die Lorr-McNair-Prozedur (Lorr, 1966b) Ähnlichkeitsindizes wie Korrelations- oder Kongruenzkoeffizienten. Dadurch wird nur die Formähnlichkeit zwischen Profilen berücksichtigt. Falls auch Niveauunterschiede berücksichtigt werden, so sollen diese nach Auffinden von Typen zu einer weiteren Aufspaltung der Typen verwendet werden. Die Verwendung des euklidischen Distanzmaßes (Lorr, Klett & McNair, 1963, S. 135), wie es z.B. von Cattell (1949) in  $r_p$  und Sawrey, Keller & Conger (1960) verwendet wird, wird abgelehnt, weil Form- und Niveauunterschiede vermengt sind und ein und dieselbe Abstand sehr unterschiedlich interpretiert

werden kann. Die Methode geht davon aus, daß ein Typ eine Klasse von Personen ist, bei der jede Person in der Klasse im Mittel jeder anderen Person in der Klasse ähnlicher ist als irgendeiner Person in irgendeiner anderen Klasse (Lorr, 1966b, S. 26). Nach Standardisierung der Variablen und Aufstellung der Korrelationsmatrix umfaßt das Verfahren die folgenden Schritte:

1. Für jedes Profil in der Korrelationsmatrix werden alle Profile mit einem Koeffizienten oberhalb einer Obergrenze  $C_1$  aufgelistet.
2. Es wird das Profil aufgesucht, mit dem die größte Anzahl anderer Profile assoziiert ist.
3. Man bildet ein Paar bestehend aus dem Profil aus Schritt 1 und dem Profil, das den höchsten mittleren Korrelationskoeffizienten mit allen Profilen der Liste aus Schritt 1 hat.
4. Es wird eine Triade gebildet, indem man zu dem Paar das Profil hinzunimmt, das den höchsten mittleren Korrelationskoeffizienten mit den Elementen der beiden zu dem Paar gehörigen Listen aufweist.
5. Aus der ganzen Matrix, mit Ausnahme der Profile aus der Triade, werden sukzessiv die Profile dem Cluster hinzugefügt, die den höchsten mittleren Korrelationskoeffizienten mit den sich schon im Cluster befindlichen Profilen haben.
6. Das Verfahren bricht ab, wenn es keine Profile mehr gibt, deren mittlerer Korrelationskoeffizient mit dem Cluster den Wert  $C_1$  übersteigt.
7. Falls sich im Cluster mehr als 5 Profile befinden und dieser Zyklus nicht der erste ist, eliminiert man alle Profile aus Clustern, die in früheren Zyklen gebildet wurden, die einen mittleren Korrelationskoeffizienten mit dem augenblicklichen Cluster haben, der größer als  $C_u$  ist.
8. Aus der Restmatrix eliminiert man alle Profile, die einen mittleren Korrelationskoeffizienten von mindestens  $C_u$  mit den Elementen jedes der bereits gebildeten Cluster haben.
9. Falls im letzten Cluster mehr als 4 Elemente sind, führt man Schritt 10 durch. Sonst wiederholt man Schritt 1 bis Schritt 8.
10. Man gibt für jedes Cluster die konstituierenden Profile, die mittleren Korrelationskoeffizienten innerhalb und zwischen den Clustern und die mittleren Korrelationskoeffizienten jedes Profils mit jedem Cluster an.

Dieses Verfahren ist mehrfach zur Identifikation von Typen verwendet worden. So von Lorr, Bishop & McNair (1965) bei psychiatrischen Patienten, von McNair & Lorr (1965) bei ambulanten psychiatrischen Patienten, von Gold-

stein & Linden (1969) bei Alkoholikern, von Suziedelis & Lorr (1973) zur Berufsdifferenzierung, von Suziedelis, Lorr & Tonesk (1974) zur Identifikation von Werturteilstypen, von Lorr, Pokorny & Klett (1973) zum Auffinden von Depressionstypen, von Tonesk, Suziedelis & Lorr (1974) zur Findung von Berufsinteressentypen bei Männern, von Youniss & Lorr (1972) zur Identifikation von Persönlichkeitstypen und von Berzius, Ross, English & Haley (1974) bei Drogensüchtigen.

Anhand der Daten einer Stichprobe von Psychotikern vergleichen Lorr & Radhakrishnan (1967) dieses Verfahren mit einem anderen Verfahren, das zunächst Clusterzentren zu bestimmen sucht durch Berechnung der Varianz der quadrierten Korrelationen jeder Person mit allen anderen (vgl. Tryon, 1958). Im weiteren Verlauf wird die *Multiple-Gruppen-Faktor-Analyse* von Guttman (1952) verwendet. Beide Verfahren liefern disjunkte Untergruppen, die sich teilweise im Umfang unterscheiden.

Die Ergebnisse einer Simulationsstudie für das oben beschriebene Verfahren werden von Suziedelis, Lorr & Tonesk (1976) diskutiert. In dieser Studie werden verschiedene Versionen des Verfahrens miteinander verglichen. So wird neben einem Satz von 96 Items auch ein daraus erhaltener Satz von 12 Summenscores betrachtet. Als Ähnlichkeitskoeffizienten werden der Produktmomentkorrelationskoeffizient und der Cohen-Koeffizient (Cohen, 1969) betrachtet. Außerdem werden verschiedene Kriterien der Typenbildung berücksichtigt. Es zeigt sich eine gewisse Überlegenheit des Verfahrens, die sich in einer besseren Replikation der Typen ausdrückt, bei Verwendung der Summenscores.

#### 4.5 Die Gruppierungsanalyse von Friedman und Rubin

In Friedman & Rubin (1967) wird eine Cluster-Analyse beschrieben, die davon ausgeht, daß n Objekte in g Gruppen zu zerlegen sind. Dazu wird im Prinzip die Menge aller Zerlegungen der n Objekte in g Gruppen betrachtet. Für jede Zerlegung wird für jede Gruppe die Matrix der Abweichungsquadrate der Elemente der Gruppe vom Gruppenschwerpunkt bestimmt. Die Summe dieser Matrizen ergibt eine Matrix W. Man sucht diejenige Zerlegung, für die W eine minimale Determinante hat.

Dieses Verfahren wird von Paykel (1971) verwendet, um Typen depressiver Patienten aufzufinden. In Paykel & Henderson (1977) erfolgt ein Vergleich dieser Methode und der gefundenen Typologie mit den Ergebnissen verschiedener anderer Cluster-Methoden.

## 4.6 Der informationstheoretische Ansatz von Wallace und Boulton

Ein Problem bei Typenanalysen, die auf Ähnlichkeitsmaßen beruhen, ist die Identifikation unterschiedlicher Typologien je nach verwendetem Ähnlichkeitsmaß. Dabei ist es meist nicht möglich, den Erfolg einer Typenbildung durch ein Maß zu messen, das unabhängig von dem typenanalytischen Prozeß ist. Deshalb versuchen Wallace & Boulton (1968) ein solches Gütemaß anzugeben, das von dem Prozeß der Typenfindung völlig unabhängig ist.

Falls an N Personen jeweils m Variablen gemessen worden sind, so läßt sich diese Information in einer ( $m \times N$ )-Matrix darstellen. Nach einer Klassenbildung kann diese Information durch folgende Angaben wiedergegeben werden:

1. Die Angabe der Klasse, zu der eine Person gehört.
2. Die durchschnittlichen Eigenschaften der Klasse.
3. Die Abweichungen jeder Person von den durchschnittlichen Eigenschaften der Klasse.

Falls die Klassen bzw. Typen klar ausgeprägt sind, so werden die Abweichungen von den Durchschnittseigenschaften klein sein und die meiste Information ist in der Angabe der Klasse enthalten, zu der eine Person gehört. In diesem Falle kann die enthaltene Information weit kürzer mitgeteilt werden, als dieses der Fall ohne Klassenbildung wäre. Sowohl im Falle diskreter als auch im Falle stetig verteilter Variablen definieren Wallace & Boulton (1968) ein Informationsmaß, das die Gesamtkodierungsinformation angibt. Es wird dann diejenige Klassifikation gesucht, die dieses Informationsmaß minimiert. Eine Anwendung dieses Verfahrens bei der Suche nach Depressionstypen beschreiben Pilowsky, Levine & Boulton (1969).

## 4.7 Die lineare Typenanalyse von Overall und Mett

Die lineare Typenanalyse von Overall & Klett (1972, S. 216-238) ist ein Verfahren, um bei Vorliegen von Intervallskalendaten die Beziehungen zwischen den beobachteten Personen und den in der Population vorhandenen reinen Typen aufzudecken. Man geht aus von der Vorstellung, daß einer beliebigen Population relativ wenige *reine Typen* zugrundeliegen. Jede Person kann eine Mischung verschiedener Typen sein, wobei sie i.a. einem speziellen reinen Typ am ähnlichsten ist. Ohne die Zahl oder das Aussehen der reinen Typen zu kennen, soll mit Hilfe der *linearen Typenanalyse* festgestellt werden, wie viele reine Typen vorliegen und wodurch sie charakterisiert sind. Dazu definiert man für jeden reinen Typ einen *Prototypvektor* bzw. ein **Profil** und

bestimmt das Ausmaß der Ähnlichkeit jeder Person mit jedem reinen Typ. Man kann dann jede Person dem Typ zuordnen, dem sie am ähnlichsten ist.

Ausgangspunkt ist eine lineare Modellgleichung

$$x = b_1 z_1 + \dots + b_r z_r + b_0 e,$$

in der  $x$  der Profilvektor des Individuums ist,  $z_1, \dots, z_r$  die Prototypprofile für  $r$  reine Typen,  $e$  ein Fehlervektor und  $b_0, b_1, \dots, b_r$  skalare Gewichte. Bei  $p$  Dimensionen und  $n$  Personen ergibt sich unter Vernachlässigung des Fehlerterms die Matrizengleichung

$$X = BZ',$$

wobei  $X$  eine  $(n \times p)$ -Matrix,  $B$  eine  $(n \times r)$ -Matrix und  $Z'$  eine  $(r \times p)$ -Matrix ist. In entsprechender Weise gibt es eine Darstellung der reinen Typen durch die Personenprofile aufgrund von

$$Z = X'A$$

wobei  $A$  eine  $(n \times r)$ -Koeffizientenmatrix ist.

Als Ähnlichkeitsmaß für zwei Profile wird das Skalarprodukt verwendet. Die Ähnlichkeitsmatrix für die Personen wird damit durch die  $(n \times n)$ -Matrix

$$Q = XX'$$

gegeben. Vor der Berechnung von  $Q$  lassen sich je nach Problemlage verschiedene Transformationen der Matrix  $X$  der Ausgangsprofile vornehmen. Man kann die Zeilenmittelwerte von den Profilscores abziehen, um Niveaugleichheit zu erzielen, man kann Zeilen- und Spaltenmittelwerte abziehen und schließlich kann man auch durch die Standardabweichungen normieren.

Die  $(n \times r)$ -Matrix

$$F = QAS$$

enthält die Ähnlichkeiten der ursprünglichen Profile mit den  $r$  reinen Typen, wobei die  $(r \times r)$ -Diagonalmatrix  $S$  für die richtige Normierung sorgt. Die Matrix der Prototypprofile ergibt sich aus

$$Z = X'AS.$$

Für die Bestimmung der gesuchten Transformationen, Prototypprofile und Gewichtskoeffizienten stehen verschiedene Verfahren zur Verfügung. In Overall & Klett (1972) werden eine *direkte Clusterrotationsanalyse* oder eine *Potenzvektorfaktorenanalyse* vorgeschlagen. Das letztere Verfahren wird in einem FORTRAN-Programm (Overall & Klett, 1972, S. 231-238) für die lineare Typenanalyse verwendet.

## 4.8 Die parametrische Mischungsanalyse von Wolfe und Fleiss

In Fleiss (1972) werden verschiedene Techniken zum Auffinden von Typen einander kritisch gegenübergestellt. Insbesondere werden die Nachteile der Q-Faktor-Analyse wie auch schon in Fleiss & Zubin (1969) und Fleiss, Lawlor, Platman & Fieve (1971) herausgearbeitet. Eine besondere Schwierigkeit bereitet die Angabe eines Verfahrens, das zwar vorhandene Typen aufdeckt, aber bei nichtvorhandenen Typen, z.B. bei einer homogenen Population, keine Typen identifiziert. Eine Gruppe heißt dabei nach Fleiss (1972) homogen, falls die gemeinsame Verteilung der Variablen multivariat normal ist, und *inhomogen*, falls die gemeinsame Verteilung eine Mischung von multivariaten Normalverteilungen ist. Vor Durchführung einer Typenanalyse kann man dann versuchen, auf Homogenität zu testen, um nur im Falle der Heterogenität nach Typen zu suchen. Als Homogenitätskriterien wurden bei diesem Ansatz, wie Fleiss referiert, entweder der Chiquadratanpassungstest auf Vorliegen einer einzigen Normalverteilung oder das Vorliegen einer unimodalen Verteilung verwendet. In Fleiss (1972) werden diese Kriterien als zu streng angesehen, da der Chiquadratatest für die speziell zu betrachtenden Alternativen nicht trennscharf genug sei und unter Umständen auf die Schiefe der tatsächlich zugrundeliegenden Verteilung und nicht auf die Heterogenität anspricht. Auch das Vorliegen nur eines Modalwertes ist kein unbedingtes Kriterium für Homogenität, da eine Mischung von zwei Normalverteilungen nur unter ganz bestimmten Voraussetzungen eine bimodale Verteilung liefert. Von Fleiss (1972) wird als verbessertes Kriterium ein auf Maximum-Likelihood-Schätzungen beruhender Chiquadrat-Test konstruiert, der die Nullhypothese einer Normalverteilung gegen die Alternativhypothese von zwei Normalverteilungen testet. Dabei ist der Test auf eine einzige Variable beschränkt, d.h. es wird nur das eindimensionale Problem betrachtet.

Eine Alternative schlägt Jones (1968) vor. Dieser bildet eine Linearkombination über die Variablen in der Weise, daß die verwendeten Gewichte dazu führen, daß die Modalwerte für die Verteilung dieser Linearkombination möglichst weit auseinanderliegen. Auf diese Weise können Typen, falls welche vorhanden sind, leichter aufgedeckt werden. Eine Kombination mit dem Verfahren von Fleiss (1972) ist naheliegend.

Einen allgemeineren Ansatz als den von Fleiss (1972) stellt die *allgemeine Mischungsanalyse* von Wolfe (1970) dar, die auch für andere Verteilungen als die multivariate Normalverteilung verwendet werden kann. Die Grundidee ist wieder eine Schätzung der Parameter der Mischverteilung mit der Maximum-Likelihood-Methode und ein darauf basierender Chiquadratstest, mit dem man die Nullhypothese, daß  $r$  Typen vorliegen, gegen die Alternativhypothese testet, daß  $r'$  Typen vorliegen, mit  $r' > r$ . Die Rechnungen sind durchgeführt für die Mischung von multivariaten Normalverteilungen und für die Mischung

von Multinomialverteilungen. Im letzteren Fall erhält man das Modell der *latenten Klassenanalyse*. Für das Normalverteilungsmodell stehen zwei FORTRAN-Programme zur Verfügung, das Programm NORMAP für Mischungen mit gemeinsamer Kovarianzmatrix und das Programm NORMIX für Mischungen mit verschiedenen Kovarianzmatrizen. Das Verfahren von Wolfe (1970) wird neben anderen Verfahren von Rosemann (1977) zur Identifikation von Schülertypen und zur Prädiktion von Schulleistungen verwendet.

## 4.9 Die Ähnlichkeitspartialisierungsmethode von Bolz

Nach Bolz (1977) beruht die von ihm vorgeschlagene Typenanalyse für Daten auf Intervallskalenniveau auf einem statistischen Clustermodell, bei dem jede Person als Zufallsziehung aus einer Population von Clustern angesehen wird. Das Modell geht von der Grundannahme aus, daß die Charakteristika einer Person eine Kombination der Charakteristika des Typs, zu dem die Person gehört, der allgemeinen Populationscharakteristika und der spezifischen Personencharakteristika sind. Der Typ, zu dem eine Person gerechnet wird, wird als *Primärtyp*, alle anderen Typen als *Sekundärtypen* bezeichnet.

Um die obige Aufspaltung der Charakteristika zu operationalisieren, wird der Begriff der *Ähnlichkeit* oder *Gemeinsamkeit* („resemblance“, abgekürzt „Res“) verwendet. Dabei wird folgende Fundamentalgleichung zugrundegelegt:

$$\text{Res}(\text{Person}) = \text{Res}(\text{Primärtyp}) + \text{Res}(\text{Sekundärtypen}) + \text{Res}(\text{Personenspezifität}).$$

Das Gesamtmaß an Ähnlichkeit, das eine Person mit einer anderen Person haben kann wird also in dieser fiktiven Weise in drei Teile aufgespalten. Der letzte Summand kann auf die Individualität der Person, den Stichprobenfehler, die Varianz des Primärtyps oder auf Meßfehler zurückgeführt werden. Eine Typenanalyse, die dieses Modell verwendet, muß versuchen, die totale Ähnlichkeit jeder Person so darzustellen, daß  $\text{Res}(\text{Primärtyp})$  gegen 1 strebt und die beiden anderen Terme gegen 0.

Als Maß für die Ähnlichkeit zweier Personen X und Y wird

$$\text{Res}(X, Y) = K/(K+D^2)$$

verwendet. Dabei ist

$$D^2 = (X_1 - Y_1)^2 + \dots + (X_m - Y_m)^2,$$

d.h. für jede der m Variablen wird die Differenz zwischen den beiden Personen gebildet und die Quadrate aufsummiert. Mit K wird der doppelte Median

einer Chi-quadratverteilung mit  $m$  Freiheitsgraden bezeichnet. Dieses Maß ist dem Maß  $r_p$  von Cattell (1949) ähnlich.

Als nächstes wird eine Formel benötigt, die es gestattet, den Anteil der Ähnlichkeit herauszupartialisieren, der auf einen speziellen Typ zurückgeht. Falls man mit  $\text{Res}(X,Y,Z)$  die Ähnlichkeit der Personen X und Y nach Extraktion der Ähnlichkeit beider Personen mit dem Typ Z bezeichnet, so wird als approximativ brauchbare Formel der Regressionsansatz

$$\text{Res}(X,Y,Z) = \text{Res}(X,Y) - \text{Res}(X,Z) \text{ Res}(Y,Z) \text{ Res}(Z,Z)^{-1}$$

vorgeschlagen.

Diejenige Person eines Typs, die mit allen anderen Personen des Typs die höchste Ähnlichkeit hat, heißt *Idealperson*. Unter Voraussetzung multivariater Normalverteilungen und bei Festsetzung geeigneter Schranken kann man definieren, was *signifikante Ähnlichkeit* heißen soll. Durch Mittelbildung über die Variablenvektoren einer Idealperson und der zu ihr signifikant ähnlichen Personen erhält man ein *Stichprobentypenzentrum*. Unter Verwendung dieses Zentrums und aller Personen, die zu ihm signifikant ähnlich sind, kann man ein neues Zentrum berechnen und so iterativ fortfahren, bis der Unterschied zwischen zwei aufeinanderfolgenden Zentren eine vorgegebene Schranke unterschreitet.

Berücksichtigt man nur signifikante Ähnlichkeiten, so bietet sich die Person mit der höchsten aufsummierten Ähnlichkeit mit allen anderen Personen als erste Idealperson eines Clusters an. Nach Auspartialisierung der Ähnlichkeit zu dieser Idealperson bzw. zu dem Typenzentrum aus der Ähnlichkeitsmatrix kann man eine zweite Idealperson lokalisieren usw. Falls bei diesem Verfahren die maximale aufsummierte Ähnlichkeit im Laufe des Prozesses an einer Stelle abrupt abnimmt, so wird dieses darauf zurückgeführt, daß alle Typen identifiziert sind. Nach dieser Festlegung der Typenzahl können dann nach dem oben beschriebenen Iterationsverfahren auf der Basis der Typenzentren Personen den Typen zugeordnet werden.

In Bolz (1977) werden Chi-quadrat-Tests empfohlen, um zu überprüfen, ob alle Typen identifiziert wurden. Um zu überprüfen, ob zwei Typen verschieden sind, wird Hotelling's  $T^2$  Statistik und für den Vergleich mehrerer Typen Wilk's Lambda vorgeschlagen.

## Literatur

- Baker, R. J. 1977. Exact distributions derived from two-way tables. *Applied Statistics*, 26, 199-206.
- Bartoszyk, G. D. & Lienert, G. A. 1978. Konfigurationsanalytische Typisierung von Verlaufskurven. *Zeitschrift für experimentelle und angewandte Psychologie*, 25, 1-9.
- Baumann, U. 1971. *Psychologische Taxometrie*. Bern: Huber.
- Baumann, U. 1973. Die Konfigurationsfrequenzanalyse, ein taxonomisches Verfahren. *Psychologische Beiträge*, 15, 153-168.
- Baumann, U., Gebert, A. & Lienert, G. A. 1973. Assoziationsstrukturanalyse depressiver Symptome. *Therapiewoche*, 23, 1-6.
- Bedeian, A. G. & Armenakis, A. A. 1977. A program for computing Fisher's exact probability test and the coefficient of association  $\lambda$  for  $n \times m$  contingency tables. *Educational and Psychological Measurement*, 37, 253-256.
- Berzius, J. I., Ross, W. F., English, G. E. & Haley, J. V. 1974. Subgroups among opiate addicts: A typological investigation. *Journal of Abnormal Psychology*, 83, 65-73.
- Bierschenk, B. & Lienert, G. A. 1977. Simple methods for clustering profiles and learning curves. *Didakometry*, No. 56, Department of Educational and Psychological Research, School of Education, Malmö.
- Bock, H. H. 1974. *Automatische Klassifikation*. Göttingen: Vandenhoeck & Ruprecht.
- Bolz, C. R. 1977. Typological theory and research. In: R. B. Cattell & R. M. Dreger (Hrsg.). *Handbook of modern personality theory*. 269-292. Washington: Hemisphere Publishing Corporation.
- Boulton, D. M. 1974. Remark on algorithm 434 (G2). *Communications of the ACM*, 17, 326.
- Burt, C. 1937. Correlations between persons. *British Journal of Psychology*, 28, 59-96.
- Burt, C. 1950. The factorial analysis of qualitative data. *The British Journal of Psychology, Statistical Section*, 3, 166-185.
- Cattell, R. B. 1943. The description of personality. Basic traits resolved into clusters. *Journal of Abnormal and Social Psychology*, 38, 476-506.
- Cattell, R. B. 1944. A note on correlation clusters and cluster search methods. *Psychometrika*, 9, 169-184.
- Cattell, R. B. 1949.  $r_p$  and other coefficients of pattern similarity. *Psychometrika*, 14, 279-298.
- Cattell, R. B. 1950. The principal culture patterns discoverable in the syntal dimensions of existing nations. *Journal of Social Psychology*, 32, 215-253.

- Cattell, R. B. 1952. The three basic factor analytic designs - their interrelations and derivatives. *Psychological Bulletin*, 49, 499-520.
- Cattell, R. B. 1957. Personality and motivation structure and measurement. New York: World Book Company.
- Cattell, R. B. 1968. Taxonomic principles for locating and using types (and the derived taxonome Computer program). In: B. Kleinmuntz (Hrsg.). *Formal representation of human judgement*. 99-149. New York: Wiley.
- Cattell, R. B. 1969. The profile similarity coefficient  $r_p$  in vocational guidance and diagnostic classification. *British Journal of Educational Psychology*, 39, 131-142.
- Cattell, R. B. & Coulter, M. A. 1966. Principles of behavioral taxonomy and the mathematical basis of the taxonome computer program. *British Journal of Mathematical and Statistical Psychology*, 19, 237-269.
- Cattell, R. B., Coulter, M. A. & Tsujioka, B. 1966. The taxonometric recognition of types and functional emergents. In: R. B. Cattell (Hrsg.). *Handbook of multivariate experimental psychology*. 288-329. Chicago: Rand McNally.
- Clark, J. A. & McQuitty, L. L. 1970. Some problems and elaborations of iterative intercolumnar correlational analysis. *Educational and Psychological Measurement*, 30, 773-784.
- Cohen, J. 1969. A profile similarity coefficient invariant over variable reflection. *Psychological Bulletin*, 71, 281-284.
- Cronbach, L.J. 1949a. „Pattern tabulation“: A statistical method for analysis of limited patterns of scores, with particular reference to the Rorschach test. *Educational and Psychological Measurement*, 9, 149-171.
- Cronbach, L. J. 1949b. Statistical methods applied to Rorschach scores: A review. *Psychological Bulletin*, 46, 393-429.
- Cronbach, L. J. 1950. Statistical methods for multiscore tests. *Journal of Clinical Psychology*, 6, 21-25.
- Cronbach, L. J. & Gleser, G. C. 1952. Similarity between persons and related problems of profile analysis. Technical Report No. 2, Bureau of Research and Service, University of Illinois.
- Cronbach, L. J. & Gleser, G. C. 1953. Assessing similarity between profiles. *Psychological Bulletin*, 50, 456-473.
- Dietsch, P. & Volk, W. 1977. Endogene Störungen. In: L. J. Pongratz & K.-H. Wewetzer (Hrsg.). *Handbuch der Psychologie*. Band 8, 1. Halbband. 255-329. Göttingen: Hogrefe.
- Dorsch, F. 1970. *Psychologisches Wörterbuch*. Hamburg: Meiner.
- DuMas, F. M. 1946. A quick method of analyzing the similarity of profiles. *Journal of Clinical Psychology*, 2, 80-83.
- DuMas, F. M. 1949. The coefficient of profile similarity. *Journal of Clinical Psychology*, 5, 123-131.
- DuMas, F. M. 1950. A note on the coefficient of profile similarity. *Journal of Clinical Psychology*, 6, 300-301.

- DuMas, F. M. 1956. Manifest structure analysis. Missoula: Montana State University Press.
- Duran, B. S. & Odell, P. L. 1974. Cluster analysis. Berlin: Springer.
- Ellson, D. G. 1955. A method for technological predictions. In: H. Quastler (Hrsg.). Information theory in psychology. 31-50. Glencoe: Free Press.
- English, H. B. & English, A. C. 1958. A comprehensive dictionary of psychological and psychoanalytical terms. New York: David McKay.
- Enke, H. 1975. Zusammenhänge zwischen Gleichverteilungs- und Unabhängigkeits-hypothesen bei qualitativen Merkmalen. Biometrische Zeitschrift, 17, 513-523.
- Eysenck, H. J. 1941. 'Type'-factors in aesthetic judgements. British Journal of Psychology, 31, 262-270.
- Feger, H. 1978. Konfliktleben und Konfliktverhalten. Bern: Huber.
- Fleiss, J. L. 1972. Classification of the depressive disorders by numerical typology. Journal of Psychiatric Research, 9, 141-153.
- Fleiss, J. L., Lawlor, W., Platman, S. R. & Fieve, R. R. 1971. On the use of inverted factor analysis for generating typologies. Journal of Abnormal Psychology, 77, 127-132.
- Fleiss, J. L. & Zubin, J. 1969. On the methods and theory of clustering. Multivariate Behavioral Research, 4, 235-250.
- Forehand, G. A. Jr. & McQuitty, L. L. 1959. Configurations of factor standings as predictors of educational achievement. Educational and Psychological Measurement, 19, 31-43.
- Freeman, G. H. & Halton, J. H. 1951. Note on an exact treatment of contingency, goodness of fit and other problems of significance. Biometrika, 38, 141-149.
- Friedman, H. P. & Rubin, J. 1967. On some invariant criteria for grouping data. Journal of the American Statistical Association, 62, 1159-1178.
- Gaier, E. L. & Lee, M. C. 1953. Pattern analysis: The configural approach to predictive measurement. Psychological Bulletin, 50, 140-149.
- Gaier, E. L., Lee, M. C. & McQuitty, L. L. 1953. Response patterns in a test of logical inference. Educational and Psychological Measurement, 13, 550-567.
- Gebert, A. 1978. Assoziationsstrukturanalyse und latent-class-model. Zeitschrift für experimentelle und angewandte Psychologie, 25, 46-54.
- Gengerelli, J. A. 1961. The analysis of mutual concurrences. Journal of Psychology, 52, 253-263.
- Gengerelli, J. A. 1963. A method for detecting subgroups in a population and specifying their membership. Journal of Psychology, 55, 457-468.
- Gibson, W. A. 1955. An extension of Anderson's solution for the latent structure equations. Psychometrika, 20, 69-73.
- Gigerenzer, G. 1977. Mathematische Methoden zur Klassifikation von Personen. In: G. Strube (Hrsg.). Binet und die Folgen. Die Psychologie des 20. Jahrhunderts. Band 5. 738-759. Zürich: Kindler.

- Gloning, M., Quatember, R. & Lienert, G. A. 1972. Konfigurationsfrequenzanalyse aphasiesspezifischer Testleistungen. Zeitschrift für Klinische Psychologie und Psychotherapie, 20, 115-122.
- Goldstein, S. G. & Linden, J. D. 1969. Multivariate classification of alcoholics by means of the MMPI. Journal of Abnormal Psychology, 74, 661-669.
- Graumann, C. F. 1960. Eigenschaften als Problem der Persönlichkeitsforschung. In: P. Lersch, F. Sander & H. Thomae (Hrsg.). Handbuch der Psychologie. Band 4. 87-154. Göttingen: Hogrefe.
- Graumann, C. F. 1966. Nichtsinnliche Bedingungen des Wahrnehmens. In: W. Metzger (Hrsg.). Handbuch der Psychologie. Band 1, 1. Halbband. 1031-1096. Göttingen: Hogrefe.
- Groffmann, K. J. & Schmidtke, A. 1977. Persönlichkeitspsychologische Grundlagen. In: L. J. Pontratz & K.-H. Wewetzer (Hrsg.). Handbuch der Psychologie. Band 8, 1. Halbband. 664-714. Göttingen: Hogrefe.
- Guttman, L. 1952. Multiple group methods for common factor analysis: Their bases, computation and interpretation. Psychometrika, 17, 209-222.
- Hancock, T. W. 1975. Remark on algorithm 434 (G2). Communications of the ACM, 18, 117-119.
- Heilmann, W. R., Lienert, G. A. & Maly, V. 1979. Prediction models in configural frequency analysis. Biometrical Journal, 21, 79-86.
- Hofstätter, P. R. 1940. über Typenanalyse. Archiv für die gesamte Psychologie, 105, 305-403.
- Hofstätter, P. R. 1955. über Ähnlichkeit. Psyche, 9, 54-80.
- Hofstätter, P. R. 1957. Psychologie. Frankfurt: Fischer.
- Hofstätter, P. R. 1971. Differentielle Psychologie. Stuttgart: Kröner.
- Hommers, W. 1978. Multinomiale und Bayes-statistische konfigurale Klassifikation. Psychologische Beiträge, 20, 267-276.
- Horn, J. L. 1961. Significance tests for use with  $r_p$  and related profile statistics. Educational and Psychological Measurement, 21, 363-370.
- Horst, P. 1954. Pattern analysis and configural scoring. Journal of Clinical Psychology, 10, 3-11.
- Howell, D. C. & Gordon, L. R. 1976. Computing the exact probability of an  $r$  by  $c$  contingency table with fixed marginal totals. Behavior Research Methods & Instrumentation, 8, 317.
- Janke, W. 1971. Klassifikation. In: R. Heiss (Hrsg.). Handbuch der Psychologie. Band 6. Psychologische Diagnostik. 901-929. Göttingen: Hogrefe.
- Jones, K. J. 1968. Problems of grouping individuals and the method of modality. Behavioral Science, 13, 496-515.
- Kainz, F. 1964. Das Denken und die Sprache. In: R. Bergius (Hrsg.). Handbuch der Psychologie. Band 1, 2. Halbband. 564-614. Göttingen: Hogrefe.

- Kaminski, G. 1964. Ordnungsstrukturen und Ordnungsprozesse. In: R. Bergius (Hrsg.). Handbuch der Psychologie. Band 1, 2. Halbband. 373-492. Göttingen: Hogrefe.
- Krauth, J. 1973. Nichtparametrische Ansätze zur Auswertung von Verlaufskurven. Biometrische Zeitschrift, 15, 557-566.
- Krauth, J. & Lienert, G. A. 1973a. Nichtparametrischer Nachweis von Syndromen durch simultane Binomialtests. Biometrische Zeitschrift, 15, 13-20.
- Krauth, J. & Lienert, G. A. 1973b. Die Konfigurationsfrequenzanalyse (KFA). Freiburg: Alber.
- Krauth, J. & Lienert, G. A. 1974. Zum Nachweis syndromgenerierender Symptominteraktionen in mehrdimensionalen Kontingenztafeln (Interaktionsstrukturanalyse). Biometrische Zeitschrift, 16, 203-211.
- Krauth, J. & Lienert, G. A. 1975. Konfigurationsfrequenzanalytische Auswertung von Verlaufskurven. In: W. H. Tack (Hrsg.). Bericht über den 29. Kongreß der Deutschen Gesellschaft für Psychologie. 402-404. Göttingen: Hogrefe.
- Krauth, J. & Lienert, G. A. 1978. Nonparametric two-sample comparison of learning curves based on orthogonal polynomials. Psychological Research, 40, 159-171.
- Krüger, H. P. 1979. Zur Anwendungsindikation von nonparametrischen Prädiktionsverfahren. Zeitschrift für Sozialpsychologie, 10, 94-104.
- Lancaster, H. O. 1951. Complex contingency tables treated by the partition of  $\chi^2$ . Journal of the Royal Statistical Society, Series B, 13, 242-249.
- Lancaster, H. O. 1969. The chi-squared distribution. New York: Wiley.
- Lange, H. J. & Vogel, T. 1965. Statistische Analyse von Symptomenkorrelationen bei Syndromen. Methods of Information in Medicine, 4, 83-89.
- Lazarsfeld, P. F. 1950. The logical and mathematical foundations of latent structure analysis. In: S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. H. Star & J. A. Clausen (Hrsg.). Measurement and prediction. Chapter 10. Princeton: Princeton University Press.
- Lee, J. W. 1977. Computing correlations with Q-sort data for McQuitty's pattern-analytic methods. Educational and Psychological Measurement, 37, 327-329.
- Lienert, G. A. 1969a. Die „Konfigurationsfrequenzanalyse“ als Klassifikationsmethode in der klinischen Psychologie. In: M. Irle (Hrsg.). Bericht über den 26. Kongreß der Deutschen Gesellschaft für Psychologie Tübingen 1968. 244-253. Göttingen: Hogrefe.
- Lienert, G. A. 1969b. Homogenität und Validität klinischer Symptomkonfigurationen. In: J. Siebenbrodt (Hrsg.). Bericht über den 2. Kongreß der Gesellschaft für Psychologie in der DDR. 130-137. Berlin: VEB Deutscher Verlag der Wissenschaften.
- Lienert, G. A. 1970. Konfigurationsfrequenzanalyse einiger Lysergsäurediöthylamid-Wirkungen. Arzneimittelforschung, 20, 912-913.
- Lienert, G. A. 1971a. Die Konfigurationsfrequenzanalyse: I. Ein neuer Weg zu Typen und Syndromen. Zeitschrift für Klinische Psychologie und Psychotherapie, 19, 99-115.

- Lienert, G. A. 1971b. Die Konfigurationsfrequenzanalyse: II. Hierarchische und agglutinierende KFA in der klinischen Psychologie. Zeitschrift für Klinische Psychologie und Psychotherapie, 19, 207-220.
- Lienert, G. A. 1971c. Die Konfigurationsfrequenzanalyse: III. Zwei- und Mehrstichproben KFA in Diagnostik und Differentialdiagnostik. Zeitschrift für Klinische Psychologie und Psychotherapie, 19, 291-300.
- Lienert, G. A. 1971d. Treatment changes in symptom pattern evaluated by configural frequency analysis. In: O. Vinar, Z. Votava & P. B. Bradley (Hrsg.). Advances in neuropsychopharmacology. 403-406. Amsterdam: North-Holland Publishing Company.
- Lienert, G. A. 1972. Die Konfigurationsfrequenzanalyse: IV. Assoziationsstruktur klinischer Skalen und Symptome. Zeitschrift für Klinische Psychologie und Psychotherapie, 20, 231-248.
- Lienert, G. A. 1973. Konfigurale Skalierung und Validierung von Fragebogen-Items. In: G. Reinert (Hrsg.). Bericht über den 27. Kongreß der Deutschen Gesellschaft für Psychologie in Kiel 1970. 609-616. Göttingen: Hogrefe.
- Lienert, G. A. & Kerekjarto, M. v. 1969. Möglichkeiten der Ex-post-Klassifizierung depressiver Symptome und Patienten mittels Faktoren- und Konfigurationsanalyse. In: H. Hippius & H. Seibach (Hrsg.). Das depressive Syndrom. 219-256. München: Urban & Schwarzenberg.
- Lienert, G. A. & Krauth, J. 1973a. Die Konfigurationsfrequenzanalyse: V. Kontingenzen- und Interaktions-Strukturanalyse multinär skalierter Merkmale. Zeitschrift für Klinische Psychologie und Psychotherapie, 21, 26-39.
- Lienert, G. A. & Krauth, J. 1973b. Die Konfigurationsfrequenzanalyse: VI. Profiländerungen und Symptomverschiebungen. Zeitschrift für Klinische Psychologie und Psychotherapie, 21, 100-109.
- Lienert, G. A. & Krauth, J. 1973c. Die Konfigurationsfrequenzanalyse: VII. Konstellations-, Konstellationsänderungs- und Profilkonstellationstypen. Zeitschrift für Klinische Psychologie und Psychotherapie, 21, 197-209.
- Lienert, G. A. & Krauth, J. 1973d. Die Konfigurationsfrequenzanalyse: VIII. Auswertung multivariater Versuchspläne. Zeitschrift für Klinische Psychologie und Psychotherapie, 21, 298-311.
- Lienert, G. A. & Krauth, J. 1974a. Die Konfigurationsfrequenzanalyse als Prädiktionsmodell in der angewandten Psychologie. In: L. H. Eckensberger & U. S. Eckensberger (Hrsg.). Bericht über den 28. Kongreß der Deutschen Gesellschaft für Psychologie in Saarbrücken 1972. Band 2. Psychologische Methodik und Mathematische Psychologie. 219-228. Göttingen: Hogrefe.
- Lienert, G. A. & Krauth, J. 1974b. Die Konfigurationsfrequenzanalyse: IX. Auswertung multivariater klinischer Untersuchungspläne (Teil 1). Zeitschrift für Klinische Psychologie und Psychotherapie, 22, 3-17.
- Lienert, G. A. & Krauth, J. 1974c. Die Konfigurationsfrequenzanalyse: IX. Auswertung multivariater klinischer Untersuchungspläne (Teil 2). Zeitschrift für Klinische Psychologie und Psychotherapie, 22, 108-121.

- Lienert, G. A. & Krauth, J. 1975a. Configural frequency analysis as a statistical tool for defining types. *Educational and Psychological Measurement*, 35, 231-238.
- Lienert, G. A. & Krauth, J. 1975b. Configural frequency analysis of repeated measurement designs: A multivariate distribution-free approach. *Journal of Multivariate Experimental Personality and Clinical Psychology*, 1, 204-206.
- Lienert, G. A. & Krauth, J. 1976. Multivariate Analysen qualitativer Daten. In: H. J. Bochnik & W. Pittrich (Hrsg.). *Multifaktorielle Probleme in der Medizin*. 161-166. Wiesbaden: Akademische Verlagsgesellschaft.
- Lienert, G. A. & Matussek, P. 1971. Der statistische Zusammenhang einiger Initial-Symptome bei unbehandelten Fällen endogener Depression. *Der Nervenarzt*, 42, 124-127.
- Lienert, G. A. & Wolfrum, C. 1979. Die Konfigurationsfrequenzanalyse: X. Therapiewirkungsbeurteilung mittels Prädiktions-KFA. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 27, 309-316.
- Ling, R. F. 1972. On the theory and construction of k-clusters. *The Computer Journal*, 15, 326-332.
- Ling, R. F. 1973. A probability theory of cluster analysis. *Journal of the American Statistical Association*, 68, 159-164.
- Ling, R. F. 1975. An exact probability distribution on the connectivity of random graphs. *Journal of Mathematical Psychology*, 12, 90-98.
- Lorr, M. 1966a. Typing and taxonomy. In: M. Lorr (Hrsg.). *Explorations in typing psychotics*. 7-13. Oxford: Pergamon.
- Lorr, M. 1966b. Approaches to typing: A critique. In: M. Lorr (Hrsg.). *Explorations in typing psychotics*. 14-29. Oxford: Pergamon.
- Lorr, M., Bishop, P. F. & McNair, D. M. 1965. Interpersonal types among psychiatric patients. *Journal of Abnormal Psychology*, 70, 468-472.
- Lorr, M., Klett, C. J. & McNair, D. M. 1963. *Syndromes of psychosis*. Oxford: Pergamon Press.
- Lorr, M., Pokorny, A. D. & Klett, C. J. 1973. Three depressive types. *Journal of Clinical Psychology*, 29, 290-294.
- Lorr, M. & Radhakrishnan, B. K. 1967. A comparison of two methods of cluster analysis. *Educational and Psychological Measurement*, 27, 47-53.
- Lubin, A. 1950. A note on „criterion analysis“. *Psychological Review*, 57, 54-57.
- Lubin, A. & Osburn, H. G. 1957. A theory of pattern analysis for the prediction of a quantitative criterion. *Psychometrika*, 22, 63-73.
- Lykken, D. T. 1956. A method of actuarial pattern analysis. *Psychological Bulletin*, 53, 102-107.
- March, D. L. 1972. Exact probabilities for R x C contingency tables (G2). *Communications of the ACM*, 15, 991-992.
- Masendorf, F. & Roeder, B. 1974. Typologisierung lernschwacher Schüler mit Hilfe der Konfigurationsfrequenzanalyse (KFA). *Psychologie in Erziehung und Unterricht*, 21, 327-334.

- Masendorf, F., Roeder, B. & Kretschmann, R. 1976. Zur differentiellen Erfassung von Schulleistungsänderungen mit Hilfe der Konfigurationsfrequenzanalyse. *Psychologie in Erziehung und Unterricht*, 23, 377-380.
- McNair, D. M. & Lorr, M. 1965. Differential typing of psychiatric outpatients. *The Psychological Record*, 15, 33-41.
- McQuitty, L. L. 1953. A statistical method for studying personality integration. In: O. H. Mowrer (Hrsg.). *Psychotherapy: Theory and research*. New York: Ronald Press.
- McQuitty, L. L. 1954a. Pattern analysis illustrated in classifying patients and normals. *Educational and Psychological Measurement*, 14, 598-604.
- McQuitty, L. L. 1954b. Theories and methods in some objective assessments of psychological well-being. *Psychological Monographs*, 68, No. 385, 1-28.
- McQuitty, L. L. 1954c. Pattern analysis - A statistical method for the study of types. In: W. E. Chalmers, M. K. Chandler, L. L. McQuitty, R. Stayner, D. E. Wray & M. Derber (Hrsg.). *Labor-management relations in Illini City, Volume II*. Champaign, Illinois: Institute of Labor and Industrial Relations. University of Illinois.
- McQuitty, L. L. 1955a. A Pattern-analytic method derived from a theory of individual differences in psychological well-being. In: S. B. Sells (Hrsg.). *Symposium on pattern analysis*. Randolph Field, Texas, Air University, USAF School of Aviation Medicine.
- McQuitty, L. L. 1955b. A method of pattern analysis for isolating typological and dimensional constructs. San Antonio, Texas: Headquarters Air Force Personal and Training Research Center, Research Bulletin TN-55-62.
- McQuitty, L. L. 1955c. Multiple classification by agreement analysis in the isolation of types of job-knowledge item-responses. University of Illinois, Contract. No: AF 33(038)-25726, November, Memorandum Report A-29.
- McQuitty, L. L. 1956. Agreement analysis: Classifying persons by predominant patterns of responses. *The British Journal of Statistical Psychology*, 9, 5-16.
- McQuitty, L. L. 1957a. Isolating predictor patterns associated with major criterion Patterns. *Educational and Psychological Measurement*, 17, 3-42.
- McQuitty, L. L. 1957b. Elementary linkage analysis for isolating orthogonal and oblique type and typal relevancies. *Educational and Psychological Measurement*, 17, 207-229.
- McQuitty, L. L. 1957c. A pattern analysis of descriptions of 'best' and 'poorest' mechanics compared with factor-analytic results. *Psychological Monographs*, 71, No. 446, 1-24.
- McQuitty, L. L. 1958. Job-knowledge scoring keys by items versus configurational analysis for assessing levels of mechanical experience. *Educational and Psychological Measurement*, 18, 661-680.
- McQuitty, L. L. 1959. Differential validity in some pattern analytic methods. In: B. M. Bass & I. A. Berg (Hrsg.). *Objective approaches to personality assessment*. Princeton, New Jersey: D. von Nostrand Company.

- McQuitty, L. L. 1960a. Hierarchical linkage analysis for the isolation of types. *Educational and Psychological Measurement*, 20, 55-67.
- McQuitty, L. L. 1960b. Hierarchical syndrome analysis. *Educational and Psychological Measurement*, 20, 293-304.
- McQuitty, L. L. 1960c. Comprehensive hierarchical analysis. *Educational and Psychological Measurement*, 20, 805-816.
- McQuitty, L. L. 1961a. Elementary factor analysis. *Psychological Reports*, 9, 71-78.
- McQuitty, L. L. 1961b. A method for selecting patterns to differentiate categories of people. *Educational and Psychological Measurement*, 21, 85-94.
- McQuitty, L. L. 1961c. Typal analysis. *Educational and Psychological Measurement*, 21, 677-696.
- McQuitty, L. L. 1961d. Item selection for configural scoring. *Educational and Psychological Measurement*, 21, 925-928.
- McQuitty, L. L. 1962. Multiple hierarchical classification of institutions and persons with reference to Union-management relations and psychological well-being. *Educational and Psychological Measurement*, 22, 513-531.
- McQuitty, L. L. 1963a. Rank order typal analysis. *Educational and Psychological Measurement*, 23, 55-61.
- McQuitty, L. L. 1963b. Best classifying every individual at every level. *Educational and Psychological Measurement*, 23, 337-345.
- McQuitty, L. L. 1964. Capabilities and improvements of linkage analysis as a clustering method. *Educational and Psychological Measurement*, 24, 441-456.
- McQuitty, L. L. 1965. A conjunction of rank order typal analysis and item selection. *Educational and Psychological Measurement*, 25, 949-961.
- McQuitty, L. L. 1966a. Multiple rank order typal analysis for the isolation of independent types. *Educational and Psychological Measurement*, 26, 3-11.
- McQuitty, L. L. 1966b. Single and multiple hierarchical classification by reciprocal pairs and rank order types. *Educational and Psychological Measurement*, 26, 253-265.
- McQuitty, L. L. 1966c. Improved hierarchical syndrome analysis of discrete and continuous data. *Educational and Psychological Measurement*, 26, 577-582.
- McQuitty, L. L. 1966d. Similarity analysis by reciprocal pairs for discrete and continuous data. *Educational and Psychological Measurement*, 26, 825-831. (Erratum: 27, 252).
- McQuitty, L. L. 1967a. A mutual development of some typological theories and pattern-analytic methods. *Educational and Psychological Measurement*, 27, 21-46.
- McQuitty, L. L. 1967b. Expansion of similarity analysis by reciprocal pairs for discrete and continuous data. *Educational and Psychological Measurement*, 27, 253-255.
- McQuitty, L. L. 1967c. A novel application of the coefficient of correlations in the isolation of both typal and dimensional constructs. *Educational and Psychological Measurement*, 27, 591-599.

- McQuitty, L. L. 1967d. Group based pattern analysis of the single individual. *Multivariate Behavioral Research*, 2, 529-536.
- McQuitty, L. L. 1968a. Improving the validity of crucial decisions in pattern analytic methods. *Educational and Psychological Measurement*, 28, 9-21.
- McQuitty, L. L. 1968b. Multiple clusters, types and dimensions from iterative intercolumnar correlational analysis. *Multivariate Behavioral Research*, 3, 465-477.
- McQuitty, L. L. 1970a. Hierarchical classification by multiple linkages. *Educational and Psychological Measurement*, 30, 3-19.
- McQuitty, L. L. 1970b. Hierarchical classification by multilevel reciprocity. *Educational and Psychological Measurement*, 30, 227-239.
- McQuitty, L. L. 1970c. Hierarchical classification by nominee-selectee analysis. *Educational and Psychological Measurement*, 30, 509-523.
- McQuitty, L. L. 1971a. Relaxed rank order typal analysis. *Educational and Psychological Measurement*, 31, 33-43.
- McQuitty, L. L. 1971b. A comparative study of some selected methods of pattern analysis. *Educational and Psychological Measurement*, 31, 607-626.
- McQuitty, L. L. 1971c. A short cut toward a submatrix containing only „disturbed“ individuals. *Educational and Psychological Measurement*, 31, 815-825.
- McQuitty, L. L. 1973. An analysis of additive variance in relation to a theory of types. *Educational and Psychological Measurement*, 33, 19-41.
- McQuitty, L. L. 1976. Comprehensive analysis of test item variance. *Educational and Psychological Measurement*, 36, 51-84.
- McQuitty, L. L., Abeles, N. & Clark, J. A. 1970. A study of the reliability of intraindividual personality structure by iterative intercolumnar correlational analysis. *Multivariate Behavioral Research*, 5, 159-175.
- McQuitty, L. L., Banks, R. G. & Frary, J. M. 1970. Submatrices of interassociations for scoring interrelatedness with matrices as an index of psychological disturbance. *Multivariate Behavioral Research*, 5, 479-488.
- McQuitty, L. L., Banks, R. G. & Frary, J. M. 1971. Interrelatedness with others as a measure of psychological stability. *Multivariate Behavioral Research*, 6, 471-483.
- McQuitty, L. L., Banks, R. G., Frary, J. M. & Aye, C. D. 1972. Selecting a submatrix likely to contain only „disturbed“ subjects. *Educational and Psychological Measurement*, 32, 53-59.
- McQuitty, L. L. & Clark, J. A. 1968a. Clusters from iterative, intercolumnar correlation analysis. *Educational and Psychological Measurement*, 28, 211-238.
- McQuitty, L. L. & Clark, J. A. 1968b. Hierarchical classification by reciprocal pairs of course selections in psychology. *Educational and Psychological Measurement*, 28, 659-689.
- McQuitty, L. L. & Frary, J. M. 1971. Reliable and valid hierarchical classification. *Educational and Psychological Measurement*, 31, 321-346.

- McQuitty, L. L. & Koch, V. L. 1975a. A method for hierarchical clustering of a matrix of a thousand by a thousand. *Educational and Psychological Measurement*, 35, 239-254.
- McQuitty, L. L. & Koch, V. L. 1975b. Highest entry hierarchical clustering. *Educational and Psychological Measurement*, 35, 751-766.
- McQuitty, L. L. & Koch, V. L. 1976. Highest column entry hierarchical clustering: A redevelopment and elaboration of elementary linkage analysis. *Educational and Psychological Measurement*, 36, 243-258.
- McQuitty, L. L., Price, L. & Clark, J. A. 1967. The problem of ties in a pattern analytic method. *Educational and Psychological Measurement*, 27, 787-796.
- Meehl, P. E. 1950. Configural scoring. *Journal of Consulting Psychology*, 14, 165-171.
- Mittenecker, E. 1960. Die quantitative Analyse der Persönlichkeit. In: P. Lersch, P. Sander & H. Thomae (Hrsg.). *Handbuch der Psychologie*. Band 4. 59-84. Göttingen: Hogrefe.
- Mombour, W. 1977. Systematik psychischer Störungen. In: L. J. Pongratz & K.-H. Wewetzer (Hrsg.). *Handbuch der Psychologie*. Band 8, 1. Halbband. 116-153. Göttingen: Hogrefe.
- Morris, J. D. 1976. A Computer program to define types by configural frequency analysis. *Behavior Research Methods & Instrumentation*, 8, 311.
- Müller, U., Ruppen, R., Baumann, U. & Angst, J. 1972. Mehrdimensionale Klassifikation des Drogenkonsums. *Archiv Psychiatrischer Nervenkrankheiten*, 216, 255-264.
- Osgood, C. E. & Suci, G. J. 1952. A measure of relation determined by both mean differente and profile information. *Psychological Bulletin*, 49, 251-262.
- Osselmann, J. 1979. Konfigurationsfrequenzanalyse von subjektiven Bedingungen depressiver Reaktionen. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 27, 116-134.
- Overall, J. E. & Klett, C. J. 1972. *Applied multivariate analysis*. New York: McGraw-Hill.
- Paykel, E. S. 1971. Classification of depressed patients: A cluster analysis derived grouping. *British Journal of Psychiatry*, 118, 275-288.
- Paykel, E. S. & Henderson, A. J. 1977. Application of cluster analysis in the classification of depression. *Neuropsychobiologie*, 3, 111-119.
- Pilowski, I., Levine, S. & Boulton, D. M. 1969. The classification of depression by numerical taxonomy. *British Journal of Psychiatry*, 115, 937-945.
- Rauchfleisch, U. 1974. Beziehungen zwischen Frustrationsreaktionen und Intelligenzfunktionen bei verwahrlosten Jugendlichen. *Psychologische Beiträge*, 16, 365-397.
- Rausch, E. 1966. Probleme der Metrik (Geometrisch-optische Täuschungen). In: W. Metzger (Hrsg.). *Handbuch der Psychologie*. Band 1, 1. Halbband. 776-865. Göttingen: Hogrefe.

- Rey, E. R., Klug, J. & Welz, R. 1978a. Die Anwendung von statistischen Methoden zur Analyse mehrdimensionaler Kontingenztafeln in der psychiatrischen Forschung. In: H. Hafner (Hrsg.). Monographien Psychiatrie, Band 17. Psychiatrische Epidemiologie. 219-222. Heidelberg: Springer.
- Rey, E. R., Klug, J. & Welz, R. 1978b. The application of Lienert's „Configuration frequency analysis“ in psychiatric epidemiology. Social Psychiatry, 13, 53-60.
- Roeder, B. 1974. Die Konfigurationsfrequenzanalyse (KFA) nach Krauth und Lienert. Ein handliches Verfahren zur Verarbeitung sozialwissenschaftlicher Daten, demonstriert an einem Beispiel. Kölner Zeitschrift für Soziologie und Sozialpsychologie, 26, 819-844.
- Roeder, B. 1976. Parameterfreier Vergleich verschiedener Konfigurationsfrequenzanalysen. Kölner Zeitschrift für Soziologie und Sozialwissenschaften, 28, 152-155.
- Roeder, B. & Masendorf, F. 1979. Differentielle Wirksamkeit von spielerischen versus übenden Lernmaterialien bei leistungsschwachen Kindern im zweiten Schuljahr. Psychologie in Erziehung und Unterricht, 26, 22-26.
- Rohracher, H. 1975. Charakterkunde. München: Urban & Schwarzenberg.
- Rosemann, B. 1977. Typologische Prädiktion und Schulleistungsprognose. In: W. H. Tack (Hrsg.). Bericht über den 30. Kongreß der Deutschen Gesellschaft für Psychologie in Regensburg 1976. Band 2, 334-335. Göttingen: Hogrefe.
- Sawrey, W. L., Keller, L. & Conger, J. J. 1960. An objective method of grouping profiles by distance functions and its relation to factor analysis. Educational and Psychological Measurement, 20, 651-673.
- Schick, C. P. 1971. Konstitutionelle Diagnostik. In: R. Heiss (Hrsg.). Handbuch der Psychologie. Band 6. Psychologische Diagnostik. 837-867. Göttingen: Hogrefe.
- Schlattmann, H. & Wildgrube, W. 1979. A FORTRAN IV program for configural frequency analysis. Educational and Psychological Measurement, 39, 673-675.
- Schott, E. & Schott, U. 1976. Zur persönlichkeitspezifischen Ant Hat-Klassifikation von Studienanfängern der Medizin mit unterschiedlichem Zulassungsmodus. (Eine Zweistichproben-Konfigurations-Frequenz-Analyse). Zeitschrift für experimentelle und angewandte Psychologie, 23, 93-128.
- Schubert, G. A. 1965. Jackson's judicial philosophy: An exploration in value analysis. American Political Science Review, 59, 941-963.
- Schulze, G. 1978. Ein Verfahren zur multivariaten Analyse der Bedingungen von Rangvariablen: Hierarchische Rangvarianzanalyse. Zeitschrift für Sozialpsychologie, 9, 129-141.
- Silbereisen, R. K., Oesterreich, R. & Leitner, K. 1977. Konfiguration der Einstellungen, Erfahrungen und Forderungen von Sozialhilfeempfängern. Zeitschrift für Klinische Psychologie und Psychotherapie, 25, 256-263.
- Sneath, P. H. A. & Sokal, R. R. 1973. Numerical taxonomy. San Francisco: Freeman.
- Stegie, R. & Wall, K. D. 1974. Tabellen für den exakten Test in  $3 \times 2$  Feldertafeln. EDV in Medizin und Biologie, 5, 73-82.
- Steinhausen, D. & Langer, K. 1977. Clusteranalyse. Berlin: de Gruyter.

- Steller, M. & Hommers, W. 1977. Zur Diagnose der Therapiemotivation durch konfigurale Klassifikation. *Diagnostica*, 23, 266-280.
- Stephenson, W. 1936a. The inverted factor technique. *British Journal of Psychology*, 26, 344-361.
- Stephenson, W. 1936b. Introduction to inverted factor analysis with some application to studies in orexis. *Journal of Educational Psychology*, 27, 353-367.
- Stephenson, W. 1950. A statistical approach to typology: The study of trait-universes. *Journal of Clinical Psychology*, 6, 26-38.
- Stephenson, W. 1952. Some observations on Q-technique. *Psychological Bulletin*, 49, 483-498.
- Stephenson, W. 1953. The study of behavior. *Q-technique and its methodology*. Chicago: University of Chicago Press.
- Strunz, K. 1960. Das Problem der Persönlichkeitstypen. In: P. Lersch, F. Sander & H. Thomae (Hrsg.). *Handbuch der Psychologie*. Band 4. 155-221. Göttingen: Hogrefe.
- Stucky, W. & Vollmar, J. 1975. Ein Verfahren zur exakten Auswertung von 2 x c Häufigkeitstabellen. *Biometrische Zeitschrift*, 17, 147-162.
- Suziedelis, A. & Lorr, M. 1973. Occupational differentiation by typological analysis. *Journal of Vocational Behavior*, 3, 367-374.
- Suziedelis, A., Lorr, M. & Tonesk, X. 1974. Patterns of personal values among men and women. *Personality & Social Psychology Bulletin*, 1, 25-27.
- Suziedelis, A., Lorr, M. & Tonesk, X. 1976. Comparison of item-level and score-level typological analysis: A Simulation study. *Multivariate Behavioral Research*, 11, 135-145.
- Tiedeman, D. V. 1955. On the study of types. In: S. B. Sells (Hrsg.). *Symposium on pattern analysis*. Randolph Field, Texas, Air University, USAF School of Aviation Medicine.
- Tonesk, X., Suziedelis, A. & Lorr, M. 1974. Vocational interest types of men-in-general. *Measurement & Evaluation in Guidance*, 7, 7978.
- Tryon, R. C. 1939. Cluster analysis. Correlation profile and orthometric factor analysis for the isolation of unities in mind and personality. Ann Arbor: Edward Brothers.
- Tryon, R. C. 1958. Cumulative communality cluster analysis. *Educational and Psychological Measurement*, 18, 3-35.
- Victor, W. 1972. Zur Klassifizierung mehrdimensionaler Kontingenztafeln. *Biometries*, 28, 427-441.
- Wallace, C. S. & Boulton, D. M. 1968. An information measure for classification. *The Computer Journal*, 11, 185-194.
- Watson, H. E. 1956. Agreement analysis. A note on Professor McQuitty's article. *The British Journal of Statistical Psychology*, 9, 17-20.
- Webster, H. A. 1952. A note on profile similarity. *Psychological Bulletin*, 49, 538-539.

- Wermuth, N. 1976. Anmerkungen zur Konfigurationsfrequenzanalyse. Zeitschrift für Klinische Psychologie und Psychotherapie, 24, 5-21.
- Wermuth, N. 1978. Zusammenhangsanalysen medizinischer Daten. Berlin: Springer.
- Werner, H. 1966. Intermodale Qualitäten (Synästhesien). In: W. Metzger (Hrsg.). Handbuch der Psychologie. Band 1, 1. Halbband. 278-303. Göttingen: Hogrefe.
- Welfe, J. H. 1970. Pattern clustering by multivariate mixture analysis. Multivariate Behavioral Research, 5, 329-350.
- Youniss, R. P. & Lorr, M. 1972. Varieties of personality style. Journal of Clinical Psychology, 28, 140-145.
- Zerssen, D. v. 1973. Methoden der Konstitutions- und Typenforschung. In: M. Thiel (Hrsg.). Enzyklopädie der geisteswissenschaftlichen Arbeitsmethoden. 9. Lieferung. 35-143. München: Oldenbourg.
- Zerssen, D. v. 1977. Konstitutionpsychologische Forschung. In: G. Strube, (Hrsg.). Binet und die Folgen. Die Psychologie des 20. Jahrhunderts. Band 5. 545-616. Zürich: Kindler.
- Zubin, J. 1934. The determination of response patterns in personality inventories. Psychological Bulletin, 31, 713.
- Zubin, J. 1936. A technique for pattern analysis. Psychological Bulletin, 33, 773.
- Zubin, J. 1937. The determination of response patterns in personality adjustment inventories. Journal of Educational Psychology, 28, 401-413.
- Zubin, J. 1938. A technique for measuring likemindedness. Journal of Abnormal and Social Psychology, 33, 508-516.

## Autoren-Register

Hinweis: Die kursivgedruckten Seitenangaben beziehen sich auf die Literaturverzeichnisse der Artikel.

- Abadie, J. 428  
Abeles, N. 455, 492  
Abrahamse, A. P. 230,  
  248, 289  
Adrain 11, 52  
Afifi, A. A. 331, 335  
Ahmann, J. S. 300, 332  
Ahmed, S. W. 323, 324,  
  332  
Ahrens 164, 200  
Ahrens, H. 67, 145  
Ahrens, H. J. 395, 409,  
  428  
Aigner, D.J. 140, 145,  
  237, 284  
Aitchison, J. 330, 332  
Aitken, C. G. G. 330,  
  332  
Aitkin, M. A. 173, 200  
Aldenderfer, M. S. 390,  
  391, 400, 404, 415, 428,  
  429  
Alexakos, C. E. 294, 332  
Altman, H. 320, 332  
Ambrosi, K. 325, 332  
Anderberg, M. R. 391,  
  394, 399, 400, 406, 415,  
  428  
Anderson, H. E. 308,  
  332, 342  
Anderson, J. A. 305,  
  314, 330, 332, 333  
Anderson, J. G. 140,  
  145  
Anderson, M. W. 326,  
  333  
Anderson, O. D. 243,  
  284  
Anderson, T. W. 21, 28,  
  36, 38, 52, 60, 67, 107,  
  141, 155, 190, 200, 271,  
  273, 276, 284, 297, 298,  
  304, 310, 333, 359, 361,  
  362, 364, 370, 374, 375,  
  377, 378, 379, 380, 382,  
  384  
Andrews, D. F. 192, 200  
Angst, J. 463, 493  
Aoyama, H. 311, 333  
Appelbaum, M. I. 165,  
  200  
Arabie, Ph. 402, 422,  
  423, 424, 425, 428, 429,  
  437  
Armenakis, A. A. 465,  
  483  
Armitage, P. 316, 333  
Arseven, E. 313, 341  
Asher, H. B. 140, 145  
Atchley, W. R. 428  
Avery, R. B. 320, 336  
Aye, C. D. 457, 492  
Backhaus, K. 208, 291  
Baton, D. W. 243, 285  
Baggaley, A. R. 308, 333  
Bagozzi, R. P. 61, 145  
Bahadur, R. R. 304, 329,  
  333  
Baitsch, H. 319, 320,  
  333  
Baker, F. 383, 384  
Baker, F. B. 401, 410,  
  411, 414, 417, 418, 419,  
  425, 428, 433  
Baker, R. J. 465, 483  
Balakrishnan, V. 311,  
  333  
Balas, E. 399, 428  
Bandemer, H. 21, 52,  
  192, 200  
Bandyopadhyay, S. 317,  
  333  
Banerjee, K. S. 304, 333  
Banfield, C. F. 404, 413,  
  428  
Bang, S. Y. 426, 438  
Banks, R. G. 457, 492  
Bard, Y. 208, 285  
Barlow, D. 243, 287  
Barnett, V. D. 271, 285  
Bartenwerfer, H. 59,  
  141  
Bartlett, M. S. 160, 194,  
  196, 200, 200, 271, 277,  
  285, 306, 333  
Bartoszyk, G. D. 470,  
  483  
Bartussek, D. 13, 52  
Basill, L. C. 404, 413,  
  428  
Batchelder, W. H. 367,  
  384  
Bauer, H. 2, 48, 52, 64,  
  89, 90, 106, 111, 127,  
  140, 141, 142, 143, 144,  
  145, 156, 201  
Bauer, R. K. 295, 319,  
  320, 333  
Bauer, W. 62, 85, 151  
Baumann, U. 441, 463,  
  465, 467, 472, 483, 493  
Beale, E. M. L. 361, 428  
Beall, G. 296, 333  
Becker, D. 416, 436  
Bedeian, A. G. 465, 483  
Behrens, W. U. 296, 334

- Ben-Bassat, M. 300, 334  
 Benini, R. 352, 384  
 Benning, R. D. 326, 333  
 Bent, D. H. 208, 229,  
   290, 320, 345, 436  
 Bentler, P. M. 26, 52,  
   60, 140, 146, 402, 438  
 Berge, C. 419, 429  
 Bernitzke, F. 60, 146  
 Berzius, J. I. 477, 483  
 Beutel, P. 229, 245, 285  
 Bhattacharya, P. K. 304,  
   334  
 Bierschenk, B. 470, 483  
 Birnbaum, A. 27, 43, 45,  
   52  
 Bishop, P. F. 476, 489  
 Bishop, Y. M. M. 25,  
   52, 118, 146  
 Blalock, A. B. 140, 146  
 Blalock, H. M. 237, 290  
 Blalock, H. M. Jr. 62,  
   63, 81, 87, 140, 146,  
   151, 164, 204  
 Blashfield, R. K. 390,  
   391, 400, 404, 414, 415,  
   428, 429, 435  
 Bledsoe, J. C. 321, 334  
 Bliss, C. I. 315, 335  
 Block, H. G. 396, 429  
 Bock, H. H. 391, 394,  
   396, 399, 403, 405, 406,  
   422, 423, 429, 441, 483  
 Bock, R. D. 21, 52, 67,  
   146, 164, 187, 188, 200,  
   201  
 Böck, G. 396, 429  
 Bolz, C. R. 481, 482,  
   483  
 Bonner, R. E. 413, 429  
 Boorman, S. A. 402,  
   411, 418, 420, 424, 425,  
   428, 429  
 Borg, J. 435  
 Borgatta, E. F. 370, 382,  
   386  
 Borgen, F. H. 299, 334  
 Bortz, J. 21, 25, 28, 52,  
   67, 146, 157, 201, 222,  
   225, 252, 253, 285  
 Boudon, R. 140, 146  
 Boulton, D. M. 465,  
   478, 483, 493, 495  
 Bowden, D. C. 208, 285  
 Bowker, A. H. 298, 334  
 Box, G. E. P. 243, 244,  
   245, 248, 249, 285  
 Boyce, D. E. 395, 397,  
   429  
 Boyd, J. P. 420, 429  
 Bradley, R. A. 330, 343  
 Brandstädter, J. 60, 146  
 Bredenkamp, J. 59, 61,  
   64, 77, 82, 93, 100, 124,  
   127, 146, 147, 159, 186,  
   200, 201  
 Breiman, L. 140, 145,  
   147  
 Breydo, M. D. 304, 345  
 Broek, K. van den 319,  
   339  
 Broffitt, B. 324, 335  
 Broffitt, J. D. 324, 326,  
   334, 346  
 Bron, C. 422, 429  
 Brown, G. W. 295, 304,  
   334  
 Brown, M. B. 245, 286,  
   396, 397, 430, 438  
 Brucker, P. 427, 429  
 Brunk, H. D. 377, 384  
 Brunswik, E. 124, 147  
 Bryan, J. G. 308, 334  
 Bryant, E. H. 428  
 Büttner, J. 399, 429  
 Bunemann, P. 420, 429  
 Bunge, M. 69, 147  
 Bunke, O. 304, 326, 334  
 Burkholder, D. L. 192, 201  
 Burnaby, T. P. 299, 334  
 Burt, C. 441, 446, 471,  
   483  
 Butther, J. N. 320, 335  
 Cacoullos, T. 295, 317,  
   334  
 Campbell, D. T. 59, 61,  
   67, 77, 84, 87, 116, 123,  
   124, 147, 186, 201  
 Campbell, J. P. 308, 333  
 Carleton, W. T. 208,  
   289  
 Carlson, J. E. 165, 205  
 Carmer, S. G. 200, 201  
 Carmone, F. J. 383, 386  
 Carroll, J. B. 406, 434  
 Carroll, J. D. 208, 271,  
   285, 406, 411, 416, 420,  
   421, 423, 424, 427, 428,  
   429  
 Carter, L. F. 62, 81, 87,  
   151, 164, 204, 237, 290  
 Cassady, J. M. 378, 379,  
   384  
 Cattell, A. K. S. 381,  
   384  
 Cattell, R. B. 308, 347,  
   381, 384, 441, 443, 446,  
   447, 470, 471, 472, 473,  
   475, 482, 483, 484  
 Cavalli-Sforza, L. L.  
   412, 431  
 Chan, L. S. 323, 334  
 Chang, J. J. 208, 285,  
   420, 429  
 Chang, P. C. 331, 335  
 Chen, M. M. 319, 339  
 Chi, P. Y. 325, 335  
 Chien, Y. T. 317, 326,  
   337  
 Chikuse, Y. 310, 335  
 Chinganda, E. F. 323,  
   335, 348  
 Cho, D. W. 320, 332  
 Chou, R. J. 321, 335  
 Christofides, N. 425,  
   429, 430  
 Claridge, G. 308, 337  
 Clark, J. A. 413, 435,  
   454, 455, 458, 484, 492,  
   493  
 Clarke, W. R. 324, 335  
 Clavelle, P. R. 320, 335  
 Cleveland, W. S. 311, 335

- Clunies-Ross, C. W. 304, 335  
 Cochran, W. G. 237, 285, 299, 315, 329, 335  
 Cohen, J. 154, 164, 165, 201, 204, 208, 225, 249, 250, 251, 255, 257, 259, 263, 264, 269, 271, 285, 402, 403, 430, 477, 484  
 Cohen, P. 164, 201, 208, 225, 249, 250, 251, 255, 257, 259, 263, 264, 269, 285  
 Cole, A. J. 423, 430  
 Coleman, J. S. 118, 147, 372, 384  
 Colonius, H. 420, 430  
 Conger, A. J. 6, 52, 53, 253, 286  
 Conger, J. J. 472, 475, 494  
 Cook, L. L. 45, 54  
 Cook, T. D. 59, 61, 77, 84, 87, 147  
 Cooley, W. W. 225, 286  
 Coolidge, F. L. 321, 335  
 Cooper, P. W. 306, 335  
 Cooper, T. 401, 435  
 Cormack, R. M. 405, 409, 410, 416, 430  
 Corsten, L. C. A. 430  
 Coulter, M. A. 441, 472, 484  
 Cournot, A. A. 352, 384  
 Cox, D. R. 25, 53, 118, 147, 313, 335  
 Cramer, E. M. 10, 53, 160, 165, 200, 200, 201, 307, 308, 335, 336  
 Cronbach, L.J. 443, 446, 471, 484  
 Cunningham, J. P. 420, 430  
 Cunningham, K. M. 414, 430  
 Curry, A. R. 307, 340  
 Dahme, B. 243, 286  
 D'Andrade, R. 411, 414, 430  
 Darlington, R. B. 6, 53, 253, 286  
 Darper, N. R. 14, 53  
 DasGupta, S. 294, 295, 297, 298, 304, 322, 326, 334, 336  
 David, F. N. 11, 53  
 Day, N. E. 299, 306, 314, 336  
 Day, W. H. E. 423, 430  
 Degerman, R. 420, 430  
 deLeeuw, J. 421, 438  
 Desu, M. M. 300, 301, 336, 338  
 Devroye, L. P. 328, 336  
 Dhrymes, Ph. 237, 286  
 Diekopf, B. A. 138, 147  
 Dietsch, P. 440, 484  
 Dillon, W. R. 295, 320, 328, 329, 330, 331, 336, 338  
 Dingman, H. F. 378, 379, 382, 383, 384, 388  
 DiPillo, P. J. 303, 336  
 Dixon, L. C. W. 427, 430  
 Dixon, W. J. 245, 286, 318, 319, 320, 336, 397, 415, 425, 430  
 Dorsch, F. 440, 484  
 Draper, N. R. 212, 229, 257, 286  
 Dudman, J. 364, 375, 387  
 Duzdzinski, M. L. 308, 336  
 DuMas, F. M. 443, 471, 472, 484, 485  
 Duncan, O. D. 60, 147, 148  
 Dunn, O. J. 302, 305, 320, 323, 334, 336, 337, 343, 349, 350  
 Dunteman, G. H. 308, 336  
 Duran, B. S. 391, 399, 430, 441, 485  
 Durbin, J. 248, 286  
 Durnovo, A. N. 304, 345  
 Dusserre, L. 319, 344  
 Dwyer, P. 381, 384  
 Eckes, Th. 391, 393, 395, 396, 405, 406, 409, 412, 430  
 Edelbrock, C. 414, 430, 431  
 Edgington, E. S. 401, 416, 431  
 Edwards, A. L. 14, 53, 164, 169, 201, 208, 212, 286  
 Edwards, A. W. F. 412, 431  
 Efron, B. 299, 336  
 Einot, J. 200, 201  
 Eisenbeis, R. A. 320, 336  
 Eisler, H. 417, 431  
 Ekehammar, B. 383, 387  
 Elfving, G. 298, 337  
 Ellison, B. E. 312, 337  
 Ellson, D. G. 472, 485  
 English, A. C. 462, 485  
 English, G. E. 477, 483  
 English, H. B. 462, 485  
 Enis, P. 298, 337  
 Enke, H. 468, 485  
 Enslein, K. 431  
 Erickson, M. T. 321, 347  
 Essler, W. K. 106, 147  
 Estabrook, G. F. 431  
 Evans, F. B. 140, 145  
 Evenson, R. C. 320, 332  
 Eye, A. von 307, 308, 327, 337, 400, 431  
 Eyman, R. K. 382, 383, 388  
 Eysenck, H. J. 136, 147, 308, 337, 441, 485  
 Ezekiel, M. 164, 201  
 Farki, A. 395, 397, 429

- Farris, J. S. 417, 431  
 Farver, T. B. 320, 337  
 Feger, H. 434, 463, 485  
 Feller, W. 50, 53  
 Fielding, A. 353, 359,  
   362, 364, 372, 374, 377,  
   379, 380, 381, 382, 384  
 Fienberg, S. E. 25, 52,  
   118, 146  
 Fieve, R. R. 480, 485  
 Fillbrandt, H. 427, 434  
 Finn, J.D. 21, 53, 156,  
   164, 200, 201, 286  
 Fischer, G. H. 6, 26, 35,  
   40, 44, 45, 53, 107, 118,  
   147, 352, 384  
 Fischer, K. 327, 337  
 Fisher, F. M. 147  
 Fisher, L. 327, 337, 434  
 Fisher, R. A. 14, 21, 53,  
   59, 147, 154, 201, 295,  
   296, 297, 300, 308, 337  
 Fisher, W. D. 399, 431  
 Flachsmeyer, J. 391, 431  
 Fleiss, J. L. 480, 485  
 Forehand, G. A. Jr. 448,  
   485  
 Formann, A. K. 43, 53,  
   60, 147, 372, 377, 380,  
   381, 383, 384  
 Fox, K. A. 164, 201  
 Frane, J. W. 208, 286  
 Frank, O. 401, 431  
 Frary, J. M. 457, 458,  
   492  
 Fraser, D. A. S. 164, 201  
 Freedman, D. A. 317,  
   337  
 Freeman, G. H. 465,  
   485  
 Friedman, H. P. 477,  
   485  
 Friendly, M. 423, 431  
 Fu, K. S. 317, 326, 337  
 Fu, King-Sun 429  
 Fuller, W. A. 286  
 Furnival, G. M. 257,  
   286, 397, 431  
 Gabriel, K. R. 173, 196,  
   200, 201, 202  
 Gadenne, V. 124, 148  
 Gaebelein, J. 165, 202  
 Gaensslen, H. 14, 53,  
   67, 148, 156, 164, 202,  
   208, 212, 222, 225, 233,  
   250, 255, 257, 278, 279,  
   286  
 Gänssler, P. 140, 141,  
   143, 148  
 Gaier, E. L. 443, 471,  
   485  
 Gal, S. 300, 334  
 Galton, F. 11, 53  
 Garnes, P. A. 200, 202  
 Ganesalingam, S. 299,  
   322, 337, 338  
 Garrett, H. E. 307, 338  
 Garside, M. J. 257, 286  
 Gartside, P. S. 227, 286  
 Gebert, A. 381, 384,  
   467, 483, 485  
 Geisser, S. 295, 298,  
   300, 301, 310, 317, 336,  
   337, 338  
 Gengerelli, J. A. 473,  
   474, 475, 48fi  
 Gentle, J. E. 427, 433  
 Gessaman, M. P. 325,  
   326, 338, 345  
 Gessaman, P. H. 325, 338  
 Gibson, W. A. 27, 40,  
   53, 60, 148, 364, 366,  
   372, 375, 380, 381, 383,  
   384, 385, 471, 485  
 Gigerenzer, G. 209, 224,  
   287, 441, 485  
 Gilbert, G. G. 320, 336  
 Gill, R. D. 383, 385  
 Gillo, M. W. 160, 200,  
   205  
 Gilula, Z. 367, 385  
 Glass, G. V. 243, 287  
 Gleser, G. C. 443, 446,  
   471, 484  
 Glick, N. 302, 311, 325,  
   329, 332, 338  
 Gloning, M. 463, 486  
 Glueck, B. C. 320, 344  
 Gnanadesikan, R. 271,  
   284, 287  
 Gocka, E. F. 202  
 Goddard, J. B. 383, 385  
 Gold, R. Z. 164, 203  
 Goldberger, A. S. 60,  
   63, 140, 145, 148, 237,  
   284  
 Goldman, R. D. 308,  
   338  
 Goldstein, M. 295, 320,  
   322, 325, 328, 329, 330,  
   331, 332, 336, 338, 342  
 Goldstein, S. G. 477,  
   486  
 Good, I. J. 367, 385  
 Goodman, L. A. 25, 43,  
   53, 54, 60, 118, 148,  
   361, 362, 372, 377, 379,  
   380, 381, 385  
 Gordon, L. R. 465, 486  
 Gordon, R. A. 256, 287  
 Gottman, J. M. 243, 287  
 Gower, J. C. 416, 431  
 Granger, C. W. J. 73,  
   148  
 Graumann, C. F. 440,  
   486  
 Graybill, F. A. 155, 156,  
   161, 164, 202, 208, 212,  
   215, 216, 233, 235, 236,  
   237, 257, 287, 290  
 Green, B. F. 60, 148  
 Green, B. F. Jr. 361,  
   362, 363, 375, 379, 381,  
   385  
 Green, P. E. 383, 386  
 Groffmann, K. J. 440,  
   486  
 Gudat, U. 243, 287  
 Guilford, J. P. 1, 54  
 Gupta, A. K. 301, 339  
 Gupta, S. S. 304, 339  
 Guseman, L. F. 311,  
   339  
 Guttman, L. 477, 486

- Haagen, K. 212, 214,  
218, 223, 261, 262, 278,  
287
- Habbema, J. D. F. 300,  
319, 320, 325, 339
- Haberman, S. J. 372,  
377, 379, 380, 386
- Hager, W. 124, 127, 148
- Hakimi, S. L. 419, 436
- Haley, J. V. 477, 483
- Halton, J. H. 465, 485
- Hambleton, R. K. 45, 54
- Harnerle, A. 36, 54, 60,  
148
- Han, C. P. 299, 339
- Hancock, T. W. 465,  
486
- Hand, H. H. 308, 339
- Hanke, B. 271, 287
- Hannan, E. J. 243, 287
- Harary, F. 419, 431
- Hardyck, C. D. 200,  
204
- Harman, H. H. 38, 54
- Harper, A. E. 308, 339
- Harper, D. 355, 360,  
382, 386
- Hart, B. J. 248, 287
- Hart, D. H. 308, 345
- Harter, H. L. 8, 54, 192,  
202, 298, 339
- Hartigan, J. A. 391, 396,  
397, 399, 400, 401, 404,  
406, 415, 424, 425, 427,  
431, 432
- Hattemer, H. 307, 339
- Haugh, L. D. 73, 151
- Hausmann, D. 427, 432
- Hay, R. A. 243, 245,  
249, 289
- Hays, D. G. 370, 382,  
386
- Hays, W. L. 14, 18, 47,  
48, 54, 163, 202, 222,  
287
- Healy, M. J. R. 307, 339
- Heck, D. L. 195, 202
- Hecker, R. 320, 339
- Hehl, F.-J. 46, 57
- Heider, F. 96, 148
- Heilmann, W. R. 468,  
486
- Heise, D. R. 63, 148
- Henderson, A. J. 477,  
493
- Henn, R. 432
- Henning, H. J. 59, 100,  
148
- Henry, N. W. 26, 27,  
38, 40, 42, 43, 55, 60,  
71, 150, 331, 342, 353,  
355, 356, 364, 366, 367,  
368, 369, 370, 371, 372,  
374, 375, 376, 378, 379,  
380, 381, 383, 386, 387,  
404, 434
- Henschke, C. I. 319,  
339
- Henschke-Mason, C.  
309, 342
- Hentschel, U. 256, 291
- Herden, W. 399, 412,  
432
- Herkner, W. H. 96, 148
- Hermans, J. 300, 319,  
320, 325, 339, 430
- Herr, D. G. 165, 202
- Herrmann, T. 4, 26, 54,  
65, 149
- Herrmann, W. M. 434
- Hersen, M. 243, 287
- Higgins, W. 308, 345
- Hild, C. 164, 202
- Hildebrandt, B. 306, 339
- Hills, M. 301, 320, 331,  
339, 340
- Hinderer, K. 140, 149
- Hoadley, B. 208, 287
- Hocking, R. R. 257,  
287, 289
- Hodapp, V. 149
- Hodson, F. R. 432
- Hoel, P. G. 310, 311,  
340
- Hoerl, A. E. 8, 54, 232,  
233, 287
- Hofstätter, P. R. 441,  
486
- Hogg, R. V. 324, 326,  
334, 346
- Holland, P. W. 25, 52,  
118, 146
- Holling, H. 255, 287
- Holtzmann, W. 243, 287
- Holzkamp, K. 1, 54
- Hommers, W. 466, 486,  
495
- Hope, K. 287
- Hopkins, C. E. 329, 335
- Hora, S. C. 301, 340
- Horn, J. L. 471, 486
- Horst, P. 38, 54, 271,  
282, 283, 284, 288, 307,  
308, 340, 472, 486
- Horst, R. 427, 432
- Horton, R. L. 156, 164,  
202
- Hosoya, Y. 73, 149
- Hotelling, H. 26, 36, 54,  
195, 202, 271, 277, 288
- Howell, D. C. 465, 486
- Huang, D. Y. 304, 339
- Huber, H. P. 243, 288
- Hubert, L.J. 397, 401,  
402, 409, 410, 411, 412,  
414, 417, 418, 419, 425,  
426, 427, 428, 432, 433,  
437
- Huberty, C. J. 295, 307,  
340
- Hudimoto, H. 326, 340
- Hudlet, R. 310, 340
- Hull, C. H. 208, 229,  
290, 320, 345, 433, 436
- Humak, K. M. S. 8, 54,  
192, 203
- Humme, U. 208, 291
- Hummell, H. J. 75, 80,  
93, 140, 149, 209, 212,  
288
- Hussy, W. 308, 337
- Hutchinson, T. P. 383,  
386

- Ikuzawa, M. 381, 386  
 Imsl 245, 257, 288  
 Isaacson, A. 375, 379,  
   386  
 Isaacson, S. L. 295, 340  
 Isii, K. 303, 340  
  
 Jackson, D. M. 253, 286  
 Jackson, D. N. 6, 52  
 Jackson, R. 300, 340  
 Jahn, J. 233, 288  
 Janke, W. 294, 295, 340,  
   440, 486  
 Jardine, N. 391, 406,  
   422, 423, 433  
 Jayachandran, K. 196,  
   204  
 Jenkins, G. H. 243, 249,  
   285  
 Jenkins, G. M. 248, 288  
 Jenkins, J. G. 208, 229,  
   290, 320, 345, 436  
 Jennings, E. 164, 205  
 Jöreskog, K. G. 2, 26,  
   36, 54, 57, 60, 63, 149,  
   209, 237, 288, 290, 418,  
   433  
 John, S. 298, 302, 340  
 Johnson, M. C. 303, 341  
 Johnson, N. L. 193, 203  
 Johnson, R. 310, 340  
 Johnson, S. C. 390, 394,  
   404, 405, 409, 411, 426,  
   433  
 Johnston, J. 60, 150,  
   164, 203, 212, 217, 245,  
   288  
 Jones, E. E. 96, 150  
 Jones, K. J. 480, 486  
  
 Kabe, D. G. 312, 349  
 Kadushin, C. 383, 386  
 Kainz, F. 440, 486  
 Kallina, H. 404, 433  
 Kaminski, G. 440, 487  
 Kanouse, D. E. 96, 150  
 Kashyap, R. L. 317, 341  
 Kasting, C. 427, 433  
  
 Katz, L. 383, 386  
 Kaufmann, A. 423, 433  
 Keeser, W. 243, 244,  
   245, 246, 288, 290  
 Keller, L. 472, 475, 494  
 Kelley, H. H. 96, 150  
 Kelly, F. J. 308, 341  
 Kendall, D. G. 132  
 Kendall, M. G. 224,  
   243, 288, 293, 294, 316,  
   326, 341, 397, 428  
 Kennard, R. W. 8, 54,  
   232, 233, 287  
 Kennedy, W. J. 427, 433  
 Kenny, D. A. 60, 140,  
   150  
 Kerbasch, J. 422, 429  
 Kerekjarto, M. v. 461,  
   488  
 Kerlinger, F. N. 14, 55,  
   225, 288  
 Kerridge, D. F. 314, 336  
 Kettenring, J. R. 271,  
   282, 283, 284, 289  
 Killough, G. G. 401,  
   435  
 Kirn, Ki Hang 426, 433  
 Kirk, R. E. 186, 203  
 Klein, L. 237, 289  
 Klein, L. R. 140, 150  
 Kleiter, E. F. 394, 419,  
   427, 434  
 Klett, C. J. 164, 204,  
   290, 475, 477, 478, 479,  
   489, 493  
 Klix, F. 434  
 Klug, J. 467, 494  
 Kmenta, J. 227, 289  
 Koch, V. L. 459, 493  
 Köpcke, W. 203  
 Koeppler, K. 59, 151  
 Koerts, J. 230, 248, 289  
 Koffler, S. L. 324, 341  
 Koller, S. 306, 339  
 Koopmans, T. C. 362,  
   364, 386  
 Kopp, B. 391, 434  
 Korte, B. 432  
  
 Kossack, C. F. 294, 295,  
   298, 341  
 Kotz, S. 193, 203  
 Kowalsky, H. J. 168,  
   203  
 Kraak, B. 65, 150  
 Krafft, O. 164, 203  
 Krauth, J. 25, 55, 462,  
   463, 464, 465, 466, 467,  
   468, 469, 470, 487, 488,  
   489  
 Kreil, C. 11, 55  
 Kretschmann, R. 464,  
   490  
 Krishnaiah, P. R. 67,  
   150, 434  
 Krishnaiah, R. 200, 203  
 Kronmal, R. A. 323,  
   330, 345, 349  
 Krüger, H. P. 468, 487  
 Kruskal, J. 406, 412, 434  
 Kruskal, J. B. 208, 289  
 Kruskal, W. H. 55, 150,  
   203  
 Krutchkoff, R. G. 208,  
   289  
 Krzanowski, W. J. 323,  
   331, 341  
 Kshirsagar, A. M. 276,  
   277, 289, 313, 321, 341  
 Kubicki, St. 434  
 Kudo, A. 298, 341  
 Küchler, M. 209, 212,  
   225, 289  
 Küffner, H. 229, 285  
 Kuiper, F. K. 434  
 Kuliback, S. 320, 331,  
   341, 378, 386  
 Kupper, L. L. 299, 342  
 Kurz, L. 317, 326, 350  
  
 Lachenbruch, P. A. 295,  
   298, 299, 301, 310, 311,  
   312, 315, 318, 321, 323,  
   324, 327, 332, 335, 341,  
   342  
 Lachin, J. M. 321, 330,  
   332, 342

- Ladd, G. W. 307, 342  
 Läuter 164, 200  
 LaFollette, W. R. 308,  
 339  
 LaMotte, L. R. 257, 289  
 Lancaster, H. O. 331,  
 342, 466, 467, 487  
 Lange, G. N. 407, 409,  
 410, 419, 434, 438  
 Lange, H. J. 461, 487  
 Langer, K. 391, 394,  
 399, 400, 405, 406, 409,  
 415, 438, 463, 494  
 Langmuir, C. R. 308,  
 347  
 Lantermann, E. D. 434  
 Laudahn, G. 434  
 Lawler, E. L. 418, 434  
 Lawley, D. N. 26, 36,  
 38, 55, 60, 150, 195, 203  
 Lawlor, W. 480, 485  
 Lazarsfeld, P. F. 4, 26,  
 27, 32, 38, 40, 42, 43,  
 55, 60, 71, 93, 150, 331,  
 342, 352, 353, 355, 356,  
 357, 359, 361, 362, 364,  
 366, 367, 368, 369, 370,  
 371, 372, 374, 375, 378,  
 379, 380, 381, 382, 383,  
 386, 387, 404, 434, 471,  
 473, 487  
 Lee, H. B. 401, 434  
 Lee, J. W. 460, 487  
 Lee, M. C. 443, 471, 485  
 Leeuw, J. de 208, 289,  
 292  
 Leitner, K. 463, 494  
 Leonhard, T. 192, 203  
 Leslie, R. N. 257, 287  
 Leton, D. A. 308, 342  
 Leuschner, D. 426, 434  
 Levine, S. 478, 493  
 Levitt, P. R. 424, 425,  
 428  
 Lewis, T. 271, 285  
 Lienert, G. A. 25, 39,  
 55, 230, 289, 417, 435,  
 460, 461, 462, 463, 464,  
 465, 466, 467, 468, 469,  
 470, 483, 486, 487, 488,  
 489  
 Lin, Y. G. 308, 343  
 Lindeman, R. H. 164,  
 203  
 Linden, J. D. 310, 347,  
 477, 486  
 Lindley, D. V. 192, 203  
 Ling, R. F. 401, 435,  
 475, 489  
 Lingoes, J. C. 401, 431  
 Linhart, H. 318, 342  
 Linn, R. L. 237, 290  
 Lippe, P. von der 352,  
 387  
 Lissitz, R. W. 309, 342  
 Littrow, J. J. 11, 55  
 Lloyd, B. B. 436  
 Lösel, F. 208, 289  
 Loeve, M. 140, 150  
 Lohmöller, J. B. 63, 150,  
 209, 271, 287, 289  
 Lohnes, P. R. 225, 286,  
 309, 321, 342  
 Lohrberg, W. 208, 291  
 Lord, F. M. 28, 36, 45,  
 55, 118, 150, 355, 360,  
 371, 382, 387  
 Lorr, M. 475, 476, 477,  
 489, 490, 495, 496  
 Love, W. 291  
 Lovell, M. C. 208, 289  
 Lowe, V. W. 322, 343  
 Lubin, A. 447, 448, 472,  
 489  
 Lumsden, J. 118, 150  
 Lykken, D. T. 448, 472,  
 489  
 MacCorquodale, K. 4,  
 26, 55  
 MacFadyen, H. W. 308,  
 342  
 MacQueen, J. B. 401,  
 434  
 MacRae, K. D. 294, 308,  
 345  
 Madansky, A. 43, 55,  
 352, 361, 364, 366, 375,  
 377, 378, 379, 387  
 Maddala, G. S. 237, 289  
 Magnusson, D. 383, 387  
 Malgady, R. G. 308, 342  
 Malinvaud, E. 60, 150,  
 230, 237, 243, 289  
 Mallinger, B. L. 308,  
 343  
 Mallows, C. L. 316, 343  
 Maly, V. 468, 486  
 Mandl, H. 271, 287  
 Mann, D. W. 397, 428  
 March, D. L. 465, 489  
 Marcus, L. F. 304, 333  
 Mardberg, B. 375, 379,  
 382, 383, 387  
 Marinell, G. 23, 55  
 Marks, S. 323, 343  
 Marquardt, D. 208, 289  
 Marsden, P. V. 140, 150  
 Marshall, A. W. 302, 343  
 Marsten, R. E. 400, 435  
 Martin, D. C. 330, 343  
 Masendorf, F. 463, 464,  
 489, 490, 494  
 Matejcek, M. 435  
 Matula, D. W. 426, 435  
 Matusita, K. 298, 311,  
 330, 343  
 Matussek, P. 463, 489  
 Maxwell, A. E. 26, 36,  
 38, 55, 60, 150  
 May, W. H. 438  
 McCabe, G. P. 318, 343  
 McCarthy, P. J. 382,  
 387  
 McCleary, R. 243, 245,  
 249, 289  
 McDonald, L. L. 322,  
 343  
 McDonald, R. P. 27, 38,  
 40, 55, 60, 150, 209,  
 289, 359, 371, 372, 388  
 McGee, V. E. 208, 289  
 McGuigan, E. J. 59,  
 109, 150

- McGuire, C. 308, 341  
 McHenry, C. E. 397,  
 435  
 McHugh, R. B. 361,  
 376, 377, 380, 383, 388  
 McIntyre, R. M. 414,  
 435  
 McKay, R. J. 310, 319,  
 320, 321, 343  
 McKeachie, W. J. 308,  
 343  
 McKeon, J. J. 271, 289  
 McLachlan, G. J. 299,  
 302, 306, 319, 322, 338,  
 343, 344  
 McLaughlin, B. 414, 431  
 McNair, D. M. 475,  
 476, 489, 490  
 McQuitty, L. L. 413,  
 435, 443, 444, 445, 446,  
 447, 448, 449, 450, 451,  
 452, 453, 454, 455, 456,  
 457, 458, 459, 471, 484,  
 485, 490, 491, 492, 493  
 Meehl, P. E. 4, 25, 26,  
 55, 444, 459, 472, 493  
 Meister, K. A. 322, 343  
 Melrose, J. P. 320, 344  
 Memon, A. Z. 298, 315,  
 344  
 Mendenhall, W. 156,  
 164, 203  
 Meo, G. de 352, 388  
 Meredith, W. 371, 388  
 Merenda, P. F. 164, 203  
 Michaelis, J. 301, 306,  
 312, 339, 344  
 Mickey, M. R. 327, 342  
 Mijares, T. A. 196, 203,  
 204  
 Mill, J. S. 109, 138, 151  
 Miller, C. R. 378, 379,  
 382, 383, 384, 388  
 Miller, R. G. 173, 200,  
 203  
 Milligan, G. W. 400,  
 409, 414, 435  
 Milliken, G. A. 208, 290  
 Mises, R. von 297, 311,  
 344  
 Mitra, S. K. 8, 56, 187,  
 204  
 Mittenecker, E. 440, 493  
 Mombour, W. 440, 493  
 Moore, D. H. 306, 330,  
 344  
 Moosbrugger, H. 2, 4,  
 6, 8, 13, 14, 16, 19, 21,  
 25, 27, 39, 42, 45, 46,  
 50, 55, 56, 57, 60, 66,  
 67, 124, 151, 154, 155,  
 156, 158, 159, 160, 162,  
 164, 166, 168, 169, 172,  
 176, 194, 195, 203, 212,  
 225, 226, 234, 290  
 Moran, M. A. 302, 319,  
 344  
 Morgan, J. A. 257, 290  
 Morris, J. D. 464, 493  
 Morrison, D. F. 195,  
 203, 290  
 Müller, H. 4, 27, 45, 46,  
 56  
 Müller, P. H. 111, 140,  
 151, 156, 203  
 Müller, R. 400, 413, 437  
 Müller, U. 463, 493  
 Müller-Merbach, H.  
 425, 436  
 Muirhead, R. J. 321, 331  
 Mulaik, S. A. 38, 56  
 Muntz, H. J. 294, 308,  
 345  
 Mustonen, S. 403, 436  
 Muthig, K. 59, 100, 148,  
 151  
 Nakache, J. P. 319, 344  
 Namboodiri, N. K. 62,  
 81, 87, 151, 164, 204,  
 237, 290  
 Narain, R. D. 315, 344  
 Narens, L. 367, 384  
 Negoita, C. V. 426, 436  
 Nenninger, P. 436  
 Nerlove, S. B. 437  
 Ness, J. W. van 327, 337  
 Neumann, J. von 248,  
 287  
 Neymark, Y. I. 304, 345  
 Nicewander, W. A. 10,  
 53, 160, 200, 201  
 Nie, N. H. 208, 229,  
 290, 320, 345, 436  
 Nie, N. N. 433  
 Nisbett, R. E. 96, 150  
 Nishi, A. 303, 345  
 Novick, M. R. 118, 250,  
 355, 360, 371, 382, 387  
 Nowick, M. R. 28, 36,  
 45, 55  
 Nygreen, G. T. 75, 151,  
 209, 291  
 Oberhofer, W. 273, 290  
 Odell, P. L. 391, 399,  
 430, 441, 485  
 Oerter, R. 271, 289  
 Oesterreich, R. 463, 494  
 Oettli, W. 432  
 Ogilvie, J. C. 114, 130  
 Okamoto, M. 298, 315,  
 344, 345  
 Oldenbürger, H.-A.  
 416, 436  
 Olivier, D. C. 411, 418,  
 420, 429  
 Olkin, I. 236, 290, 302,  
 343  
 Olson, C. L. 193, 196,  
 200, 204  
 O'Neill, T. J. 299, 345  
 Opitz, O. 391, 406, 436  
 Orth, B. 56  
 Osburn, H. G. 447, 448,  
 489  
 Osgood, C. E. 443, 471,  
 493  
 Osselmann, J. 463, 493  
 Ott, J. 330, 345  
 Overall, J. E. 164, 165,  
 204, 290, 308, 345, 478,  
 479, 493

- Padberg, M. W. 399, 428  
 Patrinos, A. N. 419, 436  
 Pawlik, K. 38, 56  
 Paykel, E. S. 477, 493  
 Pearson, E. S. 154, 204  
 Pearson, K. 1, 11, 26,  
   36, 56  
 Peay, E. R. 436  
 Pedhzur, E. J. 14, 55,  
   225, 288  
 Penfield, D. A. 324, 341  
 Perse, J. 300, 345  
 Pertier, R. 212, 214, 287  
 Petermann, F. 243, 300,  
   394, 419, 434  
 Peters, B. C. 311, 339  
 Petersen, C. R. 308, 345  
 Peterson, R. P. 310, 311,  
   340  
 Petrinovich, L. F. 200,  
   204  
 Pichot, P. 300, 345  
 Pickrel, E. W. 308, 345  
 Pierce, D. A. 73, 151,  
   243, 248, 285  
 Pillai, K. C. S. 160, 195,  
   196, 200, 204, 277, 290  
 Pilowski, I. 478, 493  
 Platman, S. R. 480, 485  
 Please, N. W. 306, 333  
 Plinke, W. 208, 291  
 Pokorny, A. D. 477, 489  
 Popper, K. R. 65, 69,  
   83, 106, 151  
 Porebski, O. R. 298,  
   312, 345  
 Porges, S. W. 321, 349  
 Power, R. P. 294, 308,  
   345  
 Pratt, J. W. 236, 290  
 Preiser, S. 59, 151  
 Prescott, E. 208, 289  
 Press, S. J. 331, 345  
 Price, L. 454, 493  
 Procter, C. 383, 386  
 Procter, C. H. 370, 388  
 Pruzansky, S. 406, 416,  
   420, 421, 429  
 Quatember, R. 463, 486  
 Quesenberry, C. P. 326,  
   345  
 Raatz, U. 59, 145  
 Rabinowitz, M. 332, 338  
 Radhakrishnan, B. K.  
   477, 489  
 Raiffa, H. 332, 345  
 Ralescu, D. A. 426, 436  
 Ralston, A. 431  
 Ramberg, J. S. 324, 326,  
   346  
 Ramsey, J. B. 227, 290  
 Rand, W. M. 400, 436  
 Randles, R. H. 324, 326,  
   334, 346  
 Rao, C. R. 8, 56, 67,  
   151, 164, 187, 194, 204,  
   212, 273, 274, 278, 290,  
   299, 307, 308, 311, 314,  
   315, 317, 321, 322, 346  
 Rao, M. M. 321, 346  
 Rao, M. R. 399, 427,  
   436  
 Rasch, D. 27, 43, 44, 56,  
   67, 151  
 Rasch, G. 21, 57, 118,  
   151  
 Rauchfleisch, U. 463,  
   493  
 Rausch, E. 440, 493  
 Rawlings, R. R. Jr. 165,  
   204  
 Reiersol, O. 362, 386  
 Remme, J. 325, 339  
 Repges, R. 326, 346  
 Revelle, W. 390, 436  
 Revenstorf, D. 38, 57,  
   243, 287, 290  
 Rey, E. R. 467, 494  
 Richards, L. E. 316,  
   327, 346  
 Richardson, S. C. 314,  
   333  
 Riffenburgh, R. H. 304,  
   335  
 Rioux, P. 307, 347  
 Rochel, H. 14, 16, 57  
 Rock, D. A. 237, 290  
 Roeder, B. 463, 464,  
   466, 489, 490, 494  
 Rogers, G. 310, 347  
 Rogers, W. H. 328, 347  
 Rohlf, F. G. 436  
 Rohlf, F. J. 423, 436,  
   437  
 Rohracher, H. 462, 294  
 Romney, A. K. 437  
 Rosch, E. 396, 436  
 Rosemann, B. 481, 494  
 Rosenthal, R. 85, 151  
 Roskam, E. E. 417, 435,  
   436  
 Ross, G. J. S. 431  
 Ross, W. F. 477, 483  
 Roßbach, H. 390, 393,  
   395, 396, 405, 406, 409,  
   412, 430  
 Rost, J. 45, 57  
 Roush, F. W. 426, 433  
 Roy, B. 437  
 Roy, S. N. 154, 195,  
   200, 204, 277, 290  
 Rubin, H. 38, 52, 107,  
   145  
 Rubin, J. 477, 485  
 Ruch, W. 46, 57  
 Rulon, P. J. 307, 308,  
   347  
 Ruppen, R. 463, 493  
 Ryan, T. A. 200, 204  
 Ryzin, J. van 325, 335,  
   437  
 Sabagh, H. 383, 388  
 Samuel, E. 317, 347  
 Sanden, A. L. van der  
   373, 383, 388  
 Sanghvi, L. D. 311, 333  
 Sarris, V. 60, 65, 66, 94,  
   151  
 Sattath, S. 416, 420, 437,  
   438  
 Sawrey, W. L. 472, 475,  
   494

- Saxena, A. K. 304, 347  
 Schaafsma, W. 295, 320,  
   322, 347  
 Schach, S. 67, 151, 164,  
   204, 212, 291  
 Schachter, J. 321, 342  
 Schäfer, T. 67, 151, 164,  
   204  
 Schäfer, Th. 212, 291  
 Scheffé, H. 67, 151, 155,  
   200, 205  
 Schenk, G. K. 435  
 Schermelleh, K. 14, 57  
 Schick, C. P. 440, 494  
 Schlattmann, H. 464, 494  
 Schmid, J. 296, 347  
 Schmidt, L. R. 308, 347  
 Schmidtke, A. 440, 486  
 Schneeweiß, H. 212,  
   219, 222, 237, 245, 291  
 Schönfeld, P. 212, 216,  
   218, 221, 233, 235, 237,  
   244, 245, 291  
 Schott, E. 466, 494  
 Schott, U. 466, 494  
 Schreiner, W. 208, 291  
 Schubert, G. A. 454, 494  
 Schubö, W. 14, 53, 67,  
   148, 156, 164, 202, 208,  
   212, 222, 225, 229, 233,  
   245, 250, 255, 256, 257,  
   278, 279, 281, 286, 291  
 Schuchard-Ficher, C.  
   208, 291  
 Schultz, J. V. 401, 410,  
   411, 418, 433, 437  
 Schulz, T. 59, 151  
 Schulze, G. 468, 494  
 Schulze, H.-H. 420, 430  
 Schweinitz, A. de 308,  
   345  
 Schwibbe, M. 416, 436  
 Searle, S. R. 21, 57, 67,  
   81, lfjl, 155, 156, 164,  
   189, 192, 196, 205  
 Seber, G. A. F. 164, 205  
 Seibel, H. D. 75, 151,  
   209, 291  
 Seifert, H.-G. 217, 218,  
   221, 223, 245, 261, 262,  
   278, 287, 291  
 Selg, H. 62, 85, 151  
 Seling, M. J. 299, 334  
 Shaffer, J. P. 160, 200,  
   205  
 Shepard, R. N. 422, 423,  
   437  
 Sheslow, D. V. 321, 347  
 Shimura, M. 439  
 Sibson, R. 391, 406, 422,  
   423, 433  
 Sievers, W. 208, 291  
 Silbereisen, R. K. 463,  
   494  
 Silverstein, A. B. 390,  
   437  
 Silvey, S. D. 256, 291  
 Simon, H. A. 63, 80,  
   152  
 Siotani, M. 303, 347  
 Sitgreaves, R. 298, 334,  
   347, 348  
 Skarabis, H. 295, 301,  
   347  
 Skene, A. M. 331, 348,  
   381, 388  
 Slater, P. 308, 346  
 Smidt, R. K. 322, 343  
 Smith, A. F. M. 192, 203  
 Smith, H. 14, 53, 212,  
   229, 257, 286  
 Smith, H. F. 295, 348  
 Smith, S. 307, 340  
 Smith, S. E. 316, 348  
 Smith, W. B. 323, 348  
 Sneath, P. H. 441, 494  
 Sneath, P. H. A. 390,  
   391, 394, 404, 416, 437  
 Sodeur, W. 416, 437  
 Sörbom, D. 36, 54, 57,  
   60, 63, 149  
 Sokal, R. R. 390, 391,  
   394, 404, 437, 441, 494  
 Sokol, L. M. 400, 435  
 Somers, R. H. 369, 383,  
   388  
 Sorum, M. 301, 348  
 Spada, H. 45, 57  
 Späth, H. 8, 57, 192,  
   205, 399, 400, 406, 413,  
   415, 437  
 Spearman, C. 59, 152  
 Spearman, C. E. 1, 26,  
   36, 37, 57  
 Spedicato, E. 427, 430  
 Spiegel, D. K. 165, 204  
 Srivastava, J. N. 323,  
   348  
 Srivastava, M. S. 311,  
   312, 348  
 Stahmann, R. F. 308,  
   348  
 Stanley, J. C. 59, 67, 77,  
   116, 123, 124, 147, 154,  
   186, 201, 205  
 Steel, R. G. D. 271, 291  
 Stegie, R. 465, 494  
 Stegmüller, W. 61, 69,  
   152  
 Steiger, J. H. 418, 438  
 Steinbrenner, K. 208,  
   229, 290, 320, 345, 436  
 Steinhausen, D. 263,  
   294, 391, 394, 399, 400,  
   405, 406, 409, 412, 415,  
   432, 438  
 Steller, M. 466, 495  
 Stephenson, W. 441,  
   460, 471, 495  
 Stevens, J. P. 193, 200,  
   205  
 Steward, D. 291  
 Steyer, R. 2, 4, 9, 10, 13,  
   14, 21, 42, 48, 50, 56,  
   57, 67, 134, 137, 140,  
   152, 154, 156, 165, 205,  
   234, 291  
 Stiefel, E. 210, 292  
 Stouffer, S. A. 383, 388  
 Streit, F. 298, 348  
 Stroebel, C. F. 320, 344  
 Stroud, T. W. F. 237,  
   291  
 Strunz, K. 440, 495

- Stuart, A. 224, 243, 288,  
     293, 341  
 Stucky, W. 465, 495  
 Student (Gosset, W. S.)  
     154, 205  
 Stute, W. 140, 141, 143,  
     148  
 Stylian, G. P. H. 295, 334  
 Subrahmaniam, K. 323,  
     348  
 Suci, G. J. 443, 471, 493  
 Süllwold, F. 59, 52  
 Sulz, K.-D. 208, 291  
 Suppes, P. 61, 78, 81,  
     152  
 Suziedelis, A. 477, 495  
 Swanson, M. R. 200,  
     201  
 Swiercinsky, D. P. 321,  
     348  
 Sydow, H. 434  
 Szegö, G. P. 427, 430  
  
 Taga, Y. 303, 340  
 Takane, J. 208, 292  
 Takane, Y. 372, 377,  
     378, 379, 380, 382, 388,  
     421, 438  
 Tallis, G. M. 317, 348  
 Tanaka, K. 439  
 Tanur, J. M. 55, 150, 203  
 Tarnai, Ch. 46, 57  
 Tatar, J. F. 257, 290  
 Tatsuoka, M. M. 23, 58,  
     67, 152, 164, 205, 291,  
     295, 308, 347, 348  
 Tautu, P. 432  
 Teichroew, D. 298, 348  
 Theil, H. 212, 216, 217,  
     222, 291  
 Theobald, C. M. 232,  
     233, 291  
 Thiele, C. 327, 337  
 Thissen, D. 237, 292  
 Thorndike, R. M. 291  
 Thurstone, L. L. 1, 26,  
     36, 58, 371, 388  
 Tiao, G. C. 243, 285  
  
 Tiedeman, D. V. 295,  
     308, 347, 348, 446, 495  
 Timäus, E. 85, 152  
 Timm, N. H. 13, 21, 58,  
     67, 152, 164, 165, 187,  
     188, 189, 191, 193, 194,  
     195, 196, 200, 205, 208,  
     273, 292  
 Timmermann, G. 419,  
     434  
 Tinhofer, G. 401, 438  
 Tintner, G. 271, 292  
 Toby, J. 383, 388  
 Tonesk, X. 477, 495  
 Torgerson, W. S. 352,  
     361, 368, 370, 371, 380,  
     381, 388  
 Toussaint, G. T. 295,  
     322, 348  
 Trampisch, H. J. 294,  
     295, 301, 330, 348, 349  
 Travers, R. M. W. 300,  
     349  
 Tryon, B. 477, 495  
 Tryon, R. C. 470, 495  
 Tsujioka, B. 472, 484  
 Tucker, L. R. 371, 388  
 Tukey, J. W. 404, 427,  
     438  
 Tversky, A. 416, 420,  
     437, 438  
 Tyron, R. C. 390, 438  
 Tzeng, O. C. S. 438  
  
 Überla, K. 38, 58, 203  
 Uematsu, T. 312, 349  
 Ullrich 50, 57  
 Urbakh, V. Y. 318, 349  
  
 Vahns, S. 96, 150  
 Van de Geer, J. B. 292  
 Van de Geer, J. P. 155,  
     205  
 Varadarajan, V. S. 317,  
     346  
 Varady, P. D. 302, 336  
 Vark, G. N. van 295,  
     320, 322, 347  
  
 Veldman, D. J. 308, 341  
 Velicer, W. F. 255, 292  
 Vetter, H. 209, 292  
 Victor, N. 301, 306,  
     325, 330, 349  
 Victor, W. 468, 495  
 Vinod, H. D. 233, 292  
 Vinogradov, B. 271, 292  
 Vogel, F. 391, 406, 416,  
     438  
 Vogel, T. 461, 487  
 Volk, W. 440, 484  
 Vollmar, J. 465, 495  
  
 Wachspress, D. P. 383,  
     386  
 Waerden, B. L. van der  
     378, 388  
 Wagner, T. J. 328, 336,  
     347  
 Wahl, P. W. 323, 349  
 Wainer, H. 208, 237,  
     292  
 Wald, A. 298, 316, 349  
 Walker, H. F. 311, 339  
 Wall, K. D. 465, 494  
 Wallace, C. S. 478, 495  
 Wallen, N. E. 308, 348  
 Walter, G. F. 321, 349  
 Wang, R. H. 303, 347  
 Wani, J. K. 312, 349  
 Ward, J. H. 164, 205,  
     405, 412, 427, 438  
 Ware, J. E. 308, 349  
 Warnock, J. K. 321, 348  
 Warren, R. 308, 338  
 Watson, G. 248, 286  
 Watson, H. E. 446, 495  
 Watts, D. G. 248, 288  
 Webster, H. 308, 349  
 Webster, H. A. 471, 495  
 Wegener, H. 320, 339  
 Weiling, F. 11, 14, 58  
 Weinberg, W. 352, 388  
 Weiner, B. 96, 150, 152  
 Weiner, J. M. 320, 349  
 Weischedel, R. 395, 397,  
     429

- Welch, B. L. 295, 297,  
305, 325, 329, 349
- Welz, R. 467, 494
- Wermuth, N. 468, 496
- Werner, H. 440, 496
- Werts, C. E. 237, 290
- Westermann, R. 124,  
127, 148
- Wexler, K. N. 420, 429
- Wheeler, L. 300, 350
- White, H. C. 424, 425,  
438
- Whitney, A. W. 327,  
350
- Wiedl, K. H. 307, 337
- Wiggins, L. 383, 388
- Wilcox, R. R. 367, 388
- Wildgrube, W. 464, 494
- Wiley, D. E. 107, 152
- Wilf, H. S. 431
- Wilkening, F. 85, 152
- Wilkening, K. 85, 152
- Wilks, S. S. 194, 205,  
277, 292
- Williams, R. G. 308, 349
- Williams, W. T. 407,  
409, 410, 419, 434, 438
- Willson, V. L. 243, 287
- Wilson, K. V. 292
- Wilson, R. W. 397, 431
- Wilson, S. 331, 345
- Winer, B. J. 21, 58, 67,  
152, 164, 186, 205
- Winter, B. B. 325, 350
- Wirsing, M. 431
- Wishart, D. 400, 415,  
423, 430, 438
- Wishart, J. 154, 193,  
194, 204
- Woinsky, M. N. 317,  
326, 350
- Wold, H. 63, 67, 152,  
153
- Wolde-Tsadik, G. 318,  
350
- Wolfe, J. H. 396, 438,  
480, 481, 496
- Wolfrum, C. 468, 489
- Wonnacott, R. J. 212,  
217, 246, 292
- Wonnacott, Th. H. 212,  
217, 246, 292
- Woodward, J. A. 402,  
438
- Wottawa, H. 16, 38, 45,  
58, 175, 205, 352,  
388
- Wright, G. H. von 71,  
152
- Wright, S. 59, 75, 79,  
153
- Wright, W. E. 415, 438
- Wüstendorfer, W. 208,  
289
- Wysotski, F. 434
- Yahil, A. 396, 438
- Yarbrough, C. 317, 350
- Yau, S. S. 316, 348
- Yeh, R. T. 426, 438
- Young, F. W. 208, 292,  
421, 438
- Youniss, R. P. 477, 496
- Yu, M. C. 319, 350
- Yule, G. U. 11, 58, 243,  
292
- Zaatar, M. K. 323, 348
- Zacks, S. 192, 205
- Zadeh, L. A. 426, 439
- Zedeck, S. 308, 350
- Zeis, C. D. 323, 348
- Zentgraf, R. 330, 349
- Zerssen, D. v. 441, 496
- Ziegler, R. 75, 80, 93,  
140, 149, 209, 212, 288
- Zielezny, M. 305, 350
- Zimmerman, D. W. 50, 58
- Zimmermann, E. 59, 73,  
153
- Zubin, J. 442, 443, 444,  
445, 449, 471, 480, 485,  
496

# Sach-Register

- Abhangigkeit**  
-, direkte kausale reglineare 88, 134ff  
-, kausale reglineare 67, 88, 93, 95,  
100f, 103ff, 123, 129, 133f, 136, 139  
-, kausale stochastische,  
s.a. kausale Abhangigkeit 78, 124  
-, logitlineare 122f, 131f  
-, nichtkausale reglineare 73, 93  
-, stochastische 74f, 133
- Ähnlichkeitsanalyse** 475ff
- allgemeines lineares Modell, s. Modell
- Alpha-Fehler, Kumulation des** 173
- Assoziationsstrukturanalyse** 466f
- Autokorrelation** 246ff
- autoregressive Prozesse 245ff
- Baumrepräsentation (von  
Proximitymatrizen)** 419ff
- Bayes-Ansatz (bei Diskriminanzanalyse)**  
300f, 313
- Bedeutsamkeit, praktische (s.a.  
Signifikanz, praktische)** 160f
- City-Block-Metrik** 395
- Clusteranalyse** 390ff, 440f
- Clusteranalytische Verfahren** 398ff  
agglomerative Verfahren 407ff, 414ff  
424
- Baumrepräsentationen** 419ff
- direkte Clusteranalyse von  
Datenmatrizen** 424f
- hierarchische Clusteranalysen** 404ff
- Mengenerlegung** 398, 406
- nicht disjunktive Clusteranalysen** 421
- subdivisive Verfahren** 412ff
- Dendrogramm** 405f, 410, 416ff
- Determinationskoeffizient**  
-, multipler 9, 11, 159  
-, multivariater 10, 160, 200
- Diskriminanzanalyse** 3f, 5, 21ff, 293ff,  
403
- , bei diskreten Daten 330ff  
-, von Zeitreihen 317f  
-, kovariate 314f  
-, multiple 308ff  
nonparametrische 325ff, 330f
- Diskriminanzfunktion** 21f, 295ff  
-, lineare 22, 295ff  
-, -, Robustheit der 323f  
-, -, Schatzung der 298, 322  
und lineare Regression 307f
- , quadratische 305ff  
-, -, Robustheit der 324
- Diskrimination**  
-, logistische 313f  
-, quadratische 305ff  
-, schrittweise 318f  
sequentialle 315ff
- Diskriminationsproblem** 293f, 318
- Distanzmae (bei Diskriminanzanalyse)**  
311
- Effekt, kausaler (s. kausalerreglinearer  
Effekt)**
- Effektgroe (s.a. Teststarke)** 262
- Erklarung, kausale** 59, 134
- Erwartungswert**  
-, bedingter 2, 7, 22, 28ff, 50ff  
-, unbedingter 31
- euklidische Metrik** 395
- Extraktionsmae (s.a. Korrelation,  
kanonische)** 278ff
- Faktor (varianzanalytischer)  
s. Variable, abhangige  
Variable, unabhangige**
- Faktorenanalyse** 36ff  
-, vs. Clusteranalyse 390, 404  
-, vs. Hauptkomponentenmethode 36  
-, vs. latente Strukturanalyse 381f
- Falsifizierung** 65, 83, 93, 97, 102, 115,  
121, 136f, 139

**Funktion**

- , lineare 29, 32, 37, 43, 45
- , logistische 30, 43
- , polynomiale 29
- , variablencharakteristische 28ff

**Gruppierungsanalyse 477f**

Guttman-Skalierung vs. latente

Strukturanalyse 370, 382

**Interaktion, varianzanalytische 80f, 83,**

96f, 99f, 114f, 120, 123

**Interaktionsstrukturanalyse 467f****Interaktionsverbot 98, 100, 116****kanonische Analyse 206f, 270ff****kausale Abhangigkeit 61ff**

kausale reglineare Abhangigkeit,  
s. Abhangigkeit

kausalerreglinearer Effekt 811, 86, 95,

100, 121, 130

-, direkter 8lf

-, totaler 86

**Klassifikationsproblem 293f**

kleinste Quadrate, Prinzip der (s.a.

Minimum-Quadrat-Prinzip) 186ff,  
210ff, 217

**Kollinearitat 249, 255ff****Konfidenzintervall 196ff****Konfigurationsfrequenzanalyse 442,**

460ff, 463, 465

agglutinierende 464f

Mehrstichproben- 464f

Pradiktions- 468

Konstanthaltung 62, 81, 108ff, 116,  
122, 137

**Kontingenzanalyse 3, 5, 23ff.****Kontrolltechniken, experimentelle**

Konstanthaltung 62, 81, 108ff, 116,  
122, 137

Parallelisierung 62, 110, 113, 115f,  
137

Randomisierung 62, 64, 98, 110,  
112ff, 122, 137

**Korrelation**

-, kanonische 270ff

Extraktionsmae 278ff

-, -, Schatzung der 276f

**Verallgemeinerung der 281ff****Redundanzmae 280f**

-, kophenetische 417

-, multiple 9, 159, 235

-, -, Bereichsschatzung der 237

-, -, Punktschatzung der 235

-, -, Teststarke bei Prufung der  
262ff

-, partielle 319

**Schein- 80**

-, semipartielles 252f, 260

-, -, Teststarke bei Prufung der  
269

**Korrelationsanalyse 5, 11ff****Kovarianz 74, 141****lateinisches Quadrat 182ff****Likelihood-Quotienten-Test 223**

-, bei Diskriminanzanalyse 316

**lokale Unabhangigkeit**

-, Axiom der 354f, 360, 377

-, Prinzip der 351f

**Maximum-Likelihood-Schatzung 190,**

376f

-, bei latenter Strukturanalyse 378ff

-, der Diskriminanzfunktion 298, 314

**Minimaxprinzip (bei Diskriminanz-  
analyse) 313f****Minimierung des erwarteten Verlustes  
(bei Diskriminanzanalyse) 310f****Minimum-Quadrat-Prinzip 210ff**

s.a. kleinste Quadrate, Prinzip der  
Minkowski-r-Metrik 395

**Modell**

-, allgemeines lineares (ALM) 155ff,  
214ff

Bereichsschatzung der Prognose  
des Erwartungswerts im  
ALM 221

Prognosen im ALM 219ff

Punktschatzung der Prognose des  
Erwartungswerts im ALM  
221

Schatzung von Parametern im  
ALM 217ff

-, allgemeines regressionsanalytisches  
(ARM) 214ff

- Zeitreihenanalyse im ARM  
243ff
- ARIMA- 245ff
  - , kausales reglineares 117
  - , klassisches korrelationsanalytisches 233ff
  - , klassisches regressionsanalytisches (KRM) 216, 219, 222
    - mit fehlerbehafteten Prädiktoren 237ff
    - statistische Tests im KRM 223ff
    - Strukturbruch 226ff
  - , logitlineares 117f, 121f, 123, 131ff
- Mehrvariablen- 133
  - , reglineares (regressiv lineares) 66, 120ff, 127, 132f, 137
- Strukturgleichungs- 60, 134
- Musterähnlichkeitsanalyse (Cattell) 470ff
- Musteranalyse (McQuitty) 442ff
  - hierarchische Syndromanalyse 449ff
- Rangordnungstypenanalyse 451f
- Typen s.a. Typ 455ff
- Übereinstimmungsanalyse 444ff
- Verbindungsanalyse 447ff
  
- Parallelisierung 62, 110, 113, 115f, 137
- Parameterschätzung
  - , bei Diskriminanzanalyse 322
  - , bei Varianzanalyse 186ff
    - Kriterium der kleinsten Quadrate 186ff
    - Maximum-Likelihood-Schätzung 186, 190
  - , im allgemeinen regressionsanalytischen Modell 217ff
- Pfданalyse 59, 62, 75, 79
- Prädiktoren, Auswahl von
  - (s.a. Regression, schrittweise, s.a. Variableauswahl) 257ff
- Austauschmethode 258f
- Methode von Cohen & Cohen 259ff
- Rückwärtsmethode 258f
- Untermengen-Methode 257
- Vorwärtsmethode 258f
- Profildistanzanalyse 472f
- Profilmodell, latentes 371
  
- Prognose (im allgemeinen regressionsanalytischen Modell) 219ff
- Proximitymaße (s.a. Ultrametrik) 392ff
  - City-Block-Metrik 395
  - Distanzfunktionen 394f
  - euklidische Metrik 395
  - Minkowski-r-Metrik 395
  - Supremum-Metrik 395
  
- Q-Analyse 441, 446, 471, 480
  
- Randomisierung 62, 64, 98, 110, 112ff, 122, 137
- Randomisierungstest (bei Clusteranalyse) 401f, 404, 416, 418
  
- Regression
  - , beschreibende lineare 209ff
  - , logistische 299
  - , multiple lineare
    - , - und Diskriminanzfunktion 307f
  - , Ridge- 230ff
  - , schrittweise 257ff, 321
- Regressionsanalyse 3f, 5, 11ff, 154f, 206ff
- Redundanzmaße (s.a. Korrelation, kanonische) 280ff
  
- Sigmaalgebra 63, 89ff
- Signifikanz, praktische 199f
  - s.a. Extraktionsmaße,
  - s.a. Redundanzmaße
- Signifikanztest
  - Definition des - 223
  - Neyman-Pearson-Test 223
    - bei autokorrierten Daten 244ff
    - bei Clusteranalyse 401f
    - bei Diskriminanzanalyse 321f
    - bei kanonischer Korrelation 277f
    - im klassischen regressionsanalytischen Modell 223ff
    - bei latenter Strukturanalyse 380f
    - bei multivariater Varianzanalyse 193ff
    - bei schrittweiser Regression 257ff
- Strukturanalyse, latente 351ff, 404
  - Modelle der - 359ff
    - latentes Distanzmodell 370

- latentes Inhaltsmodell 369
- latentes Klassenmodell 362ff
- latentes Polynommodell 367ff
- latentes Profilmodell 371f
- lokalisierter Klassenmodell 368
- testtheoretisches Modell 370f
- Parameterschätzung bei der - 372ff
- Signifikanztests bei der - 380f
  - vs. Faktorenanalyse 381f
  - vs. Guttmanskalierung 370, 382
- Supremum-Metrik 395
- Suppression 249, 253ff
  
- Taxometrie 441
- Testkriterien bei multivariater
  - Varianzanalyse 193ff
  - Hotelling-Lawley Spur-Kriterium 195
  - Pillai-Bartlett Spur-Kriterium 196
  - Roy's Eigenwert-Kriterium 195
  - Wilks Lambda-Kriterium 194
- Teststärke 261ff
  - bei Prüfung der multiplen Korrelation 262ff
  - bei Prüfung der partiellen Korrelation 269
  - bei Varianzanalysen 269
- Typ 455f
  - Agglutinations- 464
  - Anti- 462
  - , dyadischer 459
  - Einzel-Objekt 455
  - , gefleckter 458
  - Gruppen- 455
  - Gruppen-Quasi- 455
  - höherer Ordnung 459
  - Homostat- 472
  - Identitäts- 458
  - Kombinations- 469
  - Konstellations- 470
  - , linearer 460
  - Persönlichkeits- 460
  - Primär- 481
  - Profiländerungs- 468
  - , quadratischer 458
  - , realer 458
  - , reiner 456ff
  - Segregat- 472
  
- Sekundär- 481
- Testitem- 460
- Veränderungs- 464
- Verlängerungs- 458
- Typenanalyse 440ff
  - Ähnlichkeitsanalyse 475ff
  - Ähnlichkeitspartialisierungsmethode 481f
  - Gruppierungsanalyse 477f
  - Konfigurationsfrequenzanalyse 460ff
  - lineare Typenanalyse 478f
  - Mischungsanalyse 480f
  - Musterähnlichkeitsanalyse (Cattell) 470ff
  - Musteranalyse (McQuitty) 442ff
  - Profildistanzanalyse 472f
  - Übereinstimmungsanalyse (Gengerelli) 473ff
  
- Übereinstimmungsanalyse (Gengerelli) 473ff
- (McQuitty) 444ff
  - s.a. Musteranalyse (McQuitty)
- Ultrametrik 394, 397, 405, 410, 414
- Unabhängigkeit, bedingte stochastische 27f
  
- Validität
  - , externe 67, 79, 116, 123f, 131, 133
  - Populationsvalidität 124, 127ff
  - Situationsvalidität 124ff
  - Variablenvalidität 124
  - , interne 59, 61f, 66, 76ff, 138
- Variable
  - , abhängige 3, 7, 59, 61, 64, 67, 92, 113f, 156, 209
    - gekreuzte Faktoren 173ff
    - hierarchische Faktoren 179ff
    - lateinisches Quadrat 183ff
  - Dummy- 226
    - s.a. Variable, Kodier-
  - > endogene 209
  - , exogene 209
  - Indikator- 14ff, 23f, 92
    - s.a. Variable, Dummy- und Variable, Kodier-
  - , kanonische 272ff, 308
  - Kodier- 14ff

- s.a. Variable, Dummy-
- , kongenerische 26, 34, 36
- , latente 1ff, 25ff, 60, 124, 134f, 351f, 380f
- , -, Modelle mit 25ff
  - Faktorenanalyse 36ff
  - klassisches latent-additives Testmodell 45ff
  - latent class model 40ff
  - latent profile model 40ff
  - lineare traceline model 48ff
  - logistische Testmodelle 43ff
- , manifeste 2, 25, 27f, 60, 351f, 361
- Moderator- 12
- , parallele 35
- Residual- 6f, 30f, 210
  - , stochastische 6f, 11, 66, 89ff, 94, 104, 127
- Stör- 94f
  - , unabhängige 3, 59, 61, 64, 67, 92, 156, 209 s.a. Prädiktoren
  - gekreuzte Faktoren 169ff
  - hierarchische Faktoren 177ff
  - lateinisches Quadrat 182f
- Varianzanalyse 3, 5, 14ff, 154ff
  - , multivariate 153
  - , univariate 153
- Zeitreihenanalyse (im allg. regressions-analytischen Modell) 243ff